EXPERIMENTS IN PERSONALITY

Volume II

PSYCHODIAGNOSTICS
AND
PSYCHODYNAMICS

Contributors

G. CLARIDGE
Research Associate, Institute of Psychiatry

H. J. EYSENCK
Professor of Psychology, Institute of Psychiatry

J. H. G. HEWLETT
Research Associate, Institute of Psychiatry

H. C. HOLLAND
Research Associate, Institute of Psychiatry

R. W. PAYNE
Lecturer in Psychology, Institute of Psychiatry
(Now Associate Professor of Psychology, Queens University, Ontario)

P. SLATER
Lecturer in Psychological Statistics, Institute of Psychiatry

R. A. WILLETT
Lecturer in Psychology, Institute of Psychiatry

EXPERIMENTS IN PERSONALITY

Edited by

H. J. EYSENCK, Ph.D.

Professor of Psychology, University of London;
Director, Psychological Laboratories, Institute
of Psychiatry; Psychologist, Maudsley and
Bethlem Royal Hospitals

Volume II

PSYCHODIAGNOSTICS AND PSYCHODYNAMICS

S.C.E.R.T., West Bengal
Library
Calcutta
25/3, B. C. Road

London
ROUTLEDGE & KEGAN PAUL

First published 1960
by Routledge & Kegan Paul Ltd
Broadway House, Carter Lane, E.C.4

Second impression 1966

Printed in Great Britain
by Fletcher & Son Ltd
Norwich

© H. J. Eysenck 1960

No part of this book may be reproduced
in any form without permission from
the publisher, except for the quotation
of brief passages in criticism

[illegible] Ednl. [illegible]
12.3.70
2178

S.C.E.R.T., West Bengal
Date 12.3.70
Acc. No 2178

137
EYS V2

They reason theoretically, without demonstrating experimentally, and errors are the result.

MICHAEL FARADAY

CONTENTS

PART ONE: EXPERIMENTS IN PSYCHODIAGNOSTICS

1. Thought Disorder in Psychotic Patients. *R. W. Payne and J. H. G. Hewlett* *page* 3

PART TWO: EXPERIMENTS IN PSYCHODYNAMICS

2. The Excitation-Inhibition Balance in Neurotics. *Gordon Claridge* 107

The Excitation-Inhibition Balance in Normals:—

3. Measures of Learning and Conditioning. *R. A. Willett* 157
4. Measures of Perceptual Functions. *H. C. Holland* 193
5. A Factor Analysis of Selected Tests. *H. J. Eysenck* 234

PART THREE: EXPERIMENTS IN PSYCHOMETRICS

6. Factor Analysis and Some Allied Procedures. *P. Slater* 247
7. Canonical Analysis of Discriminance. *P. Slater* 256
8. Re-Examination of Some Data Collected by Hildebrand. *P. Slater* 271
Epilogue: The Place of Theory in Psychology. *H. J. Eysenck* 303
Bibliography and Author Index 316
Subject Index 329

PLATES

I. The Object-Sorting Test material 40
II. The Object Classification Test material 40
III. The Shaw Test material 41
IV. Brengelmann Picture Recognition Test material 41
V. Verbal conditioning apparatus 192
VI. Spatial conditioning apparatus: front 192
VII. Spatial conditioning apparatus: rear 192
VIII. Figural After-effect Apparatus 193
IX. Visual Tests 193

PART ONE

Experiments in Psychodiagnostics

1

THOUGHT DISORDER IN PSYCHOTIC PATIENTS

R. W. Payne and J. H. G. Hewlett

I. INTRODUCTION[1]

IT is probable that different types of thought disorder are associated with different psychiatric illnesses. This study is concerned mainly with schizophrenic thought disorder which has been described in detail by a number of psychiatrists on the basis of material collected in interviews. However, relatively few of these writers have elaborated a theory about the nature of this type of thought disorder in sufficient detail to enable objective tests to be derived from it.

(*a*) *Kretschmer's Theory*

One of the earliest psychiatrists to develop a detailed theory of schizophrenic thought disorder was Ernst Kretschmer (1936, 1951). Kretschmer accepted Bleuler's (1950) theory that a generalized mental dissociation or disintegration was the main cause of schizophrenic thought disorder, and elaborated the idea considerably. Kretschmer believed that schizophrenia was an extreme position on a fundamental constitutional variable which he called 'schizothymia-cyclothymia'. This variable was regarded as being more or less normally distributed throughout the general population, most people having some intermediate position on it. A number of characteristics, both physical and psychological, were said to be associated with this constitutional factor. According to the theory, extremely 'schizothymic' people are of leptomorphic physique, and also share a number of psychological characteristics. Extremely cyclothymic people are of pyknic physique, and share other psychological characteristics. Individuals who are abnormally 'schizothymic' tend to develop schizophrenia, while individuals who are extremely 'cyclothymic' tend to develop manic-depressive psychosis. Presumably an individual's position on this constitutional variable is largely the result of inheritance, so that only minor changes can be effected by the environment.

---

[1] The research described here was made possible by a research grant from the governors of the Maudsley and Bethlem Royal Hospitals.

This theory has a number of implications. One is that a person who is later to develop schizophrenia is from birth extremely 'schizothymic'. Even if he recovers from a schizophrenic illness, he is unlikely to have moved far along the continuum, and will always remain a 'schizothymic' personality, liable to a further illness. Another implication is that, since schizophrenia and manic-depressive psychosis are at opposite ends of the continuum, mixed psychotic states, with features of both, would be incompatible with the theory.

One of the main personality characteristics associated with 'schizothymia-cyclothymia' was held to be 'Spaltungsfähigkeit', or 'dissociation'. It too was held to be a continuous variable on which schizophrenics are extremely dissociated, and manic-depressives extremely non-dissociated. Kretschmer defined dissociation as the tendency for mental activity to occur in isolation, being uninfluenced by and isolated from other mental activity. Its opposite is a high degree of integration. An abnormal amount of dissociation could produce schizophrenic thought disorder, characterized by a 'fragmentation' of mental activity, and a lack of logical relationships between systems of ideas. Dissociation could also explain inappropriate affect, and the maintenance of delusional ideas at variance with everyday experience.

Following an early empirical study by Van der Horst (1924), Kretschmer (1928) listed a number of psychological tests which should be measures of the trait of 'dissociation'. All the measures were estimates of an individual's capacity for dissociating his attention, or performing more than one task at the same time. Some were motor, some perceptual, and some were mental tasks. If dissociation were the common factor underlying performance on these tests, they should intercorrelate significantly and yield a general factor. Payne (1955) assembled a group of typical tests of 'Spaltungsfähigkeit', including all the major tests that Kretschmer had reported on. The tests were given to 100 normal people. Seventeen 'dissociation' scores were obtained from each subject. It was found that none of these scores correlated significantly. That is to say, the matrix of correlations obtained showed only chance fluctuations from a mean value of zero. The implication is that the ability to dissociate one's attention is very specific and is entirely a function of the tasks involved. This result appears to be fatal to Kretschmer's theory. For one thing, if the 'Spaltungsfähigkeit' tests are uncorrelated, it is unlikely that they will differentiate abnormal groups in any consistent fashion. Indeed, they appear not to. Brengelmann (1954) in an extensive review of the German literature on these tests, reports that there is no consistent evidence that schizophrenics, normal 'leptosomes', normal 'pyknics' and manic-depressive psychotics obtain the mean scores on the tests demanded by the theory. In any event, Kretschmer laid great emphasis on the notion that this hypothetical factor accounts for large individual differences within the normal population. However, no such factor appears to exist.

Perhaps this result is not altogether surprising, since 'Spaltungsfähigkeit', as operationally defined by Kretschmer, is a positive ability. People who are good at dividing their attention will have an advantage in a number of cognitive situations. Even if there were evidence for a general ability of this sort, one is in the curious position of explaining a thinking deficit, schizophrenic thought disorder, as the result of an excess of some useful intellectual capacity.

(*b*) *Babcock's Theory*

Harriet Babcock (1930, 1933, 1941) advanced a different, and much simpler explanation of schizophrenic thought disorder. She suggested that in essence it is merely an extreme intellectual slowness. It is well known that people differ with respect to their speed of problem solving. Indeed Furneaux (1956) has recently demonstrated that individual differences in the speed of problem solving are an important determinant of the score obtained in a typical untimed test of general intelligence. Furthermore, intellectual speed is more or less uncorrelated with other intellectual functions. For example, fast thinkers do not necessarily make a larger number of errors. There is a very low correlation between speed and accuracy.

Babcock argued that schizophrenic thought disorder results mainly from a gross retardation of intellectual speed. Schizophrenic patients think so slowly that in a normal interview situation, they have not had time to think through the answer to one question before they have been given the next question. This makes them appear thought disordered, since one method of dealing with such a handicap is merely to answer at random. This could even explain their relative social isolation. They think too slowly to engage in normal conversation, so they tend to become socially withdrawn.

No attempt will be made to review the literature concerned with the experiments relevant to Babcock's theory, since this has been done at length elsewhere (Payne, 1960). By and large, previous investigations have tended to support Babcock's theory, in that schizophrenics have been found to be slow at problem solving. On the whole, acute or early schizophrenics are only moderately slow, while chronic or deteriorated schizophrenics are extremely slow. However, depressed patients have usually been found to be just as slow as chronic schizophrenics. These and other findings suggest the possibility that abnormal slowness of thinking is not a unitary phenomenon. It may have several causes, some people being slow because their minds work slowly, some because they are unable to concentrate, being constantly distracted, and others because they are thinking about a different problem which they believe to be relevant.

(*c*) *The Theory of 'Concreteness'*

Two entirely different theories of schizophrenic thought disorder have one thing in common. Both suggest that it is essentially due to an abnormality of concept formation of some sort. However, they disagree about the nature of this abnormality.

Kurt Goldstein and several other writers (Bolles and Goldstein, 1938; Goldstein, 1939, 1946; Hanfmann and Kasanin, 1937, 1942; Kasanin, 1946; Kasanin and Hanfmann, 1938a, 1938b) have argued that schizophrenics are abnormally 'concrete'. That is to say, they are unable to perform inductive reasoning since they are unable to make an abstract generalization. A number of experiments have been carried out to investigate this hypothesis. These will not be reviewed in detail, as this has been done elsewhere (Payne, 1960; Payne, Matussek and George, 1959). Many of these experiments make use of either the Goldstein-Scheerer (1941) sorting tests, or the Vigotsky (1934) test to assess 'concreteness'. The main fault of the studies which report schizophrenics to be abnormally 'concrete', lies in the criteria of 'concreteness' used. For example, Goldstein and Scheerer (1941) label a large number of responses as 'concrete', including a failure to sort at all, a failure to 'shift' from one method of sorting to another, a failure to give an adequate verbal account of the sorting, and several unusual ways of sorting the material. Furthermore, if the test instructions are followed exactly, individuals who are initially 'concrete' are often not allowed to continue to sort the material spontaneously, in order to see whether later they will produce a more adequate method of sorting. Instead they are given a series of 'control experiments' aimed at assisting them with the task. Similarly an unusually slow performance on the Vigotsky tests would be labelled 'concrete' if the Hanfmann-Kasanin (1937) scoring method is used, as would an unusual generalization.

More recent work (Fisher, 1950; Rashkis, Cushman and Landis, 1946; Rashkis, 1947; Rapaport *et al.*, 1945: McGaughran, 1954, 1957; McGaughran and Moran, 1956; Fey, 1951) has consistently suggested that, on sorting tests, schizophrenics tend to produce unusual generalizations when compared with normals. However, they can produce just as many generalizations as normal people, and it seems unreasonable to label this behaviour 'concrete', thus identifying it with the inability to generalize at all.

The other main method of investigating 'concreteness' in schizophrenics, has been by the use of the 'proverbs' test. Studies by Wegrocki (1940), Benjamin (1946), Becker (1956) and Gorham (1956) have produced consistent results. Schizophrenics give more 'concrete' interpretations to proverbs than do normals, whether the usual method of administration is used, or whether a multiple choice form of the test is used. In view of the results obtained with sorting tests, these results cannot be regarded as unambiguous. There is no doubt that schizophrenics tend to define words and interpret proverbs peculiarly. However, this is not necessarily due to an inability to generalize. It is possible that unusual interpretations are often arbitrarily rated as 'concrete'. It is also possible that, just as they tend to use words peculiarly, schizophrenics tend to interpret test instructions peculiarly. They may interpret the instructions as allowing the use of apt 'concrete' illustrations of the meaning the proverb has for them, as well as allowing a general explanation of the principle illustrated. It is not certain that in these studies the schizophrenics were encouraged to go on talking. Some abstract generalization might ultimately have been given, even if this were preceded by apparently 'concrete' statements, or apparently irrelevant remarks.

Previous studies then, are at best ambiguous. There is no conclusive evidence that schizophrenics are more 'concrete' than normal, although there is considerable evidence that the generalizations they make tend to be unusual.

(*d*) *Cameron's Theory of 'Overinclusion'*

Norman Cameron (1938a, 1938b, 1939a, 1939b, 1944, 1947; Cameron and Magaret, 1949, 1950, 1951) has suggested that schizophrenic thought disorder is due to a quite different abnormality of concept formation. He believes that schizophrenics' concepts are over-generalized. Schizophrenics are unable to maintain the normal conceptual boundaries, and incorporate into their concepts elements, some of them personal, which are merely associated with the concept, but are not an essential part of it. Cameron used the term 'overinclusion' to describe this abnormality. Cameron reported that in working on the Vigotsky test, and a sentence completion test, schizophrenics were unable to preserve the 'conceptual boundaries' of the task. In solving a problem, the schizophrenics '. . . included such a variety of categories at one time, that the specific problems became too extensive and too complex for a solution to be reached' (Cameron, 1939a).

A surprising number of experiments have been carried out to investigate this theory. These have been reviewed elsewhere (Payne, 1960; Payne, Matussek and George, 1959). The results obtained have consistently supported the theory. For example, Moran (1953) and Epstein (1953) developed very similar tests, consisting of a list of words printed on a sheet of paper. Following each stimulus word was a number of response words, including neologisms, and the word 'none'. The subjects were asked to underline each response word which they regarded as an essential part of the concept denoted by the stimulus word. Individuals whose concepts were abnormally overinclusive were expected to underline more words. Both investigators found that schizophrenics underlined significantly more words than normals.

White (1949), testing a matched schizoph.enic and normal group, asked his subjects to group fifteen cards with a word printed on each in any way they liked. The schizophrenics tended to form very large, vague categories, forming concepts such as 'suspicion' or 'having to do with God'.

Chapman (1956) and Chapman and Taylor (1957) report a series of interesting experiments which have also confirmed the theory of overinclusion. They made use of card sorting tests of different types, and found that, when asked to sort according to a specific concept (e.g. 'fruit'), schizophrenics tended to include in this category similar cards (e.g. cards depicting vegetables) but not completely dissimilar cards. Normals did not do this. In another experiment, the subjects were presented with cards containing four pictures, some of which illustrated a concept. The subjects were asked to sort the cards according to a concept, but to disregard all the pictures on the cards except the picture in the lower right-hand corner. As predicted, the schizophrenics were influenced by these irrelevant 'distractor' items, whereas the normals were not. There was no evidence of 'concreteness' or inability to sort at all, since the number of errors made varied directly with the number of distractor items relevant to the sorting categories, there being no significant difference between normals and schizophrenics when distractor items of this sort were not present.

It must be concluded that Cameron's theory of overinclusion has received strong support from the experiments so far carried out.

(*e*) *A Reformulation of the Theory of Overinclusion*

Payne, Matussek and George (1959) have suggested that it is possible to reformulate Cameron's theory of overinclusion in a slightly more general way so that a number of predictions follow from it. Concept formation can be regarded as largely the result of discrimination learning. When a child first hears a word in a certain context, the word is associated with the entire situation (stimulus compound). As the word is heard again and again, only certain aspects of the stimulus compound are reinforced. Gradually the extraneous elements cease to evoke the response (the word) having become 'inhibited' through lack of 'reinforcement'. This 'inhibition' is in some sense an active process, as it suppresses a response which was formerly evoked by the stimulus. 'Overinclusive thinking' may be the result of a disorder of the process whereby 'inhibition' is built up to circumscribe and define the learned response (the word or 'concept'). In short, it could be an extreme degree of 'stimulus generalization'.

The same theory can be expressed in different terms. All purposeful behaviour depends for its success on the fact that some stimuli are 'attended to' and some other stimuli are ignored. It is a well-known fact that when concentrating on one task, normal people are quite unaware of most stimuli irrelevant to the task. It is as if some 'filter mechanism' cuts out or inhibits the stimuli, both internal and external, which are irrelevant to the task in hand, to allow the most efficient 'processing' of incoming information. Overinclusive thinking might be only one aspect of a general breakdown of this 'filter' mechanism.

II. THE PRELIMINARY INVESTIGATION OF THOUGHT DISORDER IN SCHIZOPHRENICS

While many studies had been carried out to assess separately the various abnormalities described, no large-scale study had been conducted to compare measures of slowness, concreteness and overinclusion on the same schizophrenic and normal populations in order to determine which type of dysfunction is most characteristic of schizophrenia.

Payne, Matussek and George (1959) administered a large battery of tests, including measures of slowness, concreteness and overinclusion to eighteen schizophrenic and sixteen neurotic in-patients. The groups were carefully matched for age, Mill Hill Vocabulary score (Raven, 1948), Progressive Matrices score (Raven, 1948) and Nufferno 'Level' score (Furneaux, 1956), both groups being almost exactly of average mean intelligence on all three tests.

Most of the tests given have been used in the present study and will be described in detail later. In order to test Babcock's theory that abnormal intellectual slowness is the basis of schizophrenic thought disorder, three tests of intellectual speed were given, Nufferno speed tests A1 (individual, unstressed), A2 (individual stressed) and B1 (individual, stressed) (Furneaux, 1956).

The schizophrenics were slower than the neurotics, but on only one test did the mean difference achieve the 5% level of statistical significance (on a 'one tail' test). A Nufferno 'speed-level discrepancy' score also failed to differentiate the groups significantly. It was predicted that if schizophrenics were abnormally slow, they would tend to compensate for this by working at their maximum speed all the time. Thus, when their normal working rate (as measured by the 'unstressed' speed test—A1) was compared with their maximum rate (as measured by 'stressed' speed test A2) they should show little improvement in speed as compared with the neurotics. While this measure of 'stress gain' differentiated the groups as expected (i.e. schizophrenics improved less when asked to work as fast as possible), the difference was only statistically significant at the 5% level by a 'one tail' test. It was also predicted that, if schizophrenics were abnormally slow thinkers, they would tend to compensate for this by being more persistent. However, a measure of persistence consisting of the two longest times spent during the Nufferno Level test on items which were later abandoned or solved incorrectly, failed to differentiate the groups significantly. It was concluded that intellectual slowness is not an important abnormality, fundamental to thought disorder in schizophrenia, as Babcock's theory suggests. It was suggested that intellectual slowness in schizophrenia might be a secondary by-product of 'overinclusion'. Because they perceive 'irrelevant' aspects of the test as being necessary for a solution (e.g. their personal associations to the letters in the Nufferno letter-series items), schizophrenics have more data to think about. They must consider a larger number of possible solutions, and this would result in a longer average solution time per item, with a greater variability of solution times. This would not necessarily lead to an increase in errors, as solutions based on irrelevant aspects of the material might still be rejected if they did not 'fit'.

Babcock believed that abnormal slowness in schizophrenia is fundamental, and affects simple motor speed tests as well as more complex intellectual tests. To test this theory, four of Babcock and Levy's (1940) simple motor speed tests were administered. These were the speed of writing 'the United States of America', the speed of writing one's own name, the speed of writing a sentence from dictation, and the speed of performing the Babcock-Levy (1940) 'substitution' test. While the schizophrenics were slightly slower on all four measures, none of the differences reached statistical significance. Again, Babcock's theory that slowness is of fundamental importance, was not supported.

Five measures of 'concreteness' were obtained from the two groups. These were ratings obtained in the traditional way from the Goldstein-Scheerer (1941) Colour-Form and Object-Sorting tests, and Benjamin's (1946) Proverbs test, and two scores derived from a modified form of Feldman and Drasgow's (1951) test for concept formation. Only one of these scores differentiated the groups significantly. This was the rating of 'concreteness' obtained from the Goldstein-Scheerer Object-Sorting test. The schizophrenics were significantly more 'concrete'. It was concluded that the heterogeneous nature of the test material of the object-sorting test favoured the production of unusual responses in the schizophrenic group, and that for this reason they were labelled more 'concrete' when the Goldstein-Scheerer test administration and scoring criteria were followed. Goldstein and Scheerer (1941) have identified 'concreteness' with 'rigidity'. Two measures of 'rigidity' were obtained from the modified Feldman-Drasgow test. These were based on the number of times the two alternative methods of classifying the objects on each card were perceived. Only when a one-minute time limit per card was enforced did the rigidity score differentiate the groups significantly, and it was pointed out that this result was almost certainly due to the tendency among the schizophrenics to produce unusual groupings, which were not scored as correct. It was concluded that there was no evidence of abnormal concreteness in the schizophrenic group.

Thirteen different measures of 'overinclusion' were obtained from the Benjamin Proverbs test, the Goldstein-Scheerer Object-Sorting test, an 'Object Classification Test', the Epstein (1953) test for 'overinclusion', and the Leiter Partington (1950) 'Pathways' test. All these scores will be described in detail later. Seven of these scores differentiated the groups as predicted, at well beyond the 1% level of significance. Only four scores failed to produce significant differences, although the mean differences were in the expected direction.

The Klein and Krech (1952) tactile 'figural after effect' test was given to the two groups. There was evidence that the schizophrenics developed weaker figural after effects, and might be regarded as developing less 'reactive inhibition'. While this result was not regarded as unambiguous, it was consistent with the hypothesis that 'overinclusion', which was thought to be due to a defect of some central 'filtering' mechanism which inhibits irrelevant stimuli, is associated with a general inability to develop cortical 'inhibition'.

This study suggested that 'overinclusion' may be a fundamental cause of schizophrenic thought disorder. The present study was designed to follow up these results.

III. THE PURPOSE OF THE PRESENT STUDY

The studies previously mentioned, and the initial study (Payne, Matussek and George, 1959) are all consistent with the hypothesis that 'overinclusion' is an abnormality fundamental to schizophrenic thought disorder. However, all these studies have one weakness. It has not been demonstrated that abnormal 'overinclusion' is specific to schizophrenia. It might be equally frequent in depressive patients, and thus might better be regarded as a general characteristic of psychosis.

A study by Eysenck (1952a, 1952b) suggests that this is a strong possibility. Eysenck (1952a), in attempting to find evidence for Kretschmer's typology 'schizothymia-cyclothymia', tested a group of a hundred normal soldiers, fifty schizophrenic patients, and fifty manic-depressive patients. They were given a large battery of tests, most of which were selected because it was thought that they would differentiate between the two psychotic groups. Twenty of the tests which differentiated best were factor analysed separately for the normal and the psychotic groups. It was demonstrated that only one factor was necessary to account for the differences observed between the three groups. This factor was labelled 'psychoticism'. On the tests of 'psychoticism', normals had the lowest scores, schizophrenics tended to have intermediate scores, while manic-depressives tended to have the most extreme scores. It had been expected that a second factor, independent of 'psychoticism', but differentiating the two psychotic groups would be obtained. However, no such factor could be demonstrated, and the results were consistent with the theory that in the population investigated, the depressives differed from the schizophrenics only in that they were more 'psychotic'.

Two later studies (Eysenck, Granger and Brengelmann, 1957; Eysenck, S.B.G., 1955) investigated further tests of 'psychoticism'. However, in these studies, the two psychotic groups were not analysed separately, schizophrenics and manic-depressives being combined. In both researches, a number of tests successfully differentiated between a combined psychotic group, and a normal and a neurotic group.

The aims of the present investigation can be summarized as follows:

(1) To determine whether 'overinclusion' is an abnormality specific to schizophrenia, or whether it is a general characteristic of psychotic patients.

(2) To determine whether or not neurotic patients suffer from any degree of thought disorder. It is generally believed that neurosis does not produce thought disorder of the kinds described, hence the use of neurotic patients as a control group in the preliminary investigation.

(3) To determine whether or not tests of 'overinclusion' intercorrelate significantly, and thus measure some factor in common. While the preliminary study suggested that they differentiate schizophrenics from neurotics as predicted, it is necessary to demonstrate that they have something in common before we can conclude that they all measure the same basic abnormality. It is further necessary to demonstrate that the common factor shared by these tests (as measured by a factor score) also differentiates schizophrenics from the other groups.

(4) To determine whether general psycho-motor slowness is more characteristic of depressed patients than of schizophrenics. It has been mentioned that although Babcock thought psycho-motor slowness to be fundamental to schizophrenia, several studies suggest that depressed patients as a group are considerably slower than acute schizophrenic patients. The evidence so far available suggests that there might be two unrelated causes of slowness in psychotic patients. It is possible that depressives are extremely slow at all types of task, motor, perceptual and mental, possibly because they are so distracted by their depressive ideas that they are generally retarded. Slowness in acute schizophrenics on the other hand, might be confined to intellectual tasks. It is possible that they are slow at problem solving, merely because their overinclusive thinking forces them to consider an unduly large number of possibilities. If this general formulation were correct, we would expect to find in a combined population of normals, neurotics and psychotics, that tests of 'overinclusion' would be unrelated to tests of simple motor and per-

ceptual speed. Tests of reasoning speed on the other hand, would be related to both tests of overinclusion, and tests of motor and perceptual slowness, since slowness of reasoning can be caused either by general retardation or by 'overinclusion'.

(5) To investigate the relationship between the tests of 'psychoticism' used by Eysenck and his co-workers, and measures of 'overinclusive' thinking and slowness.

(6) To replicate the previous finding that schizophrenics are not abnormally 'concrete'.

IV. THE GROUPS TESTED

The subjects tested consisted of twenty normal people, twenty neurotic patients, twenty depressive patients and twenty schizophrenics. These groups were matched for age, pre-illness intellectual level (as assessed by a vocabulary test), socio-economic status, and educational attainment.

The twenty normal subjects were selected carefully to match the abnormal groups as closely as possible. They were approached individually, and were drawn from a very wide range of sources. Their backgrounds were very varied, and their occupations ranged from 'home help' to physician. Some of the occupations sampled were counter hand, carpenter, plumber, garage hand, engineering student, typist, secretary, motor mechanic, house-painter, shop assistant supervisor and general clerk. This wide range of occupation was necessary in order to match them with the patients, who were equally heterogeneous. It had initially been intended to match the groups for sex as well as the other variables. However, it proved so difficult to match the groups on five independent variables, that sex was allowed to go uncontrolled, as it was thought to be the least important factor. In the normal group, there were nine males and eleven females.

The twenty schizophrenic patients consisted of ten males and ten females. The group was composed mainly of 'acute' or 'early' schizophrenics. Only one was regarded as chronic, and the remaining patients were not judged to have deteriorated markedly by the psychiatrists in charge. On the other hand, only typical, clear cut cases were selected. A number of these patients had been admitted previously with a similar illness. All were inpatients of the Maudsley Hospital, except for two cases who were inpatients of Cane Hill Hospital. Most were tested within two weeks to one month after admission and all were judged still to be psychotic at the time of testing. These subjects were, on average, considerably more co-operative than a random sample of acute schizophrenic patients. This was because the lengthy testing programme (two days) required a considerable degree of co-operation. A number of subjects were excluded because they were unable to co-operate sufficiently. However, several quite disturbed patients were persuaded to take the entire battery. On average, the testing required four two-hour sessions, although some slower patients required longer than this. So far as possible, all the subjects were tested before they had begun treatment, and before they had been given any drugs (apart from mild night sedation). This proved a difficult condition to fulfil, as a large proportion of the psychotic inpatient population in the hospitals involved, were receiving drugs soon after admission. Only two of the schizophrenics tested were having drugs at the time of testing (apart from mild night sedation). Both were having 'Largactil' (50 mg. three times a day). No attempt was made to select any particular sub-type of schizophrenia. A number of the patients selected were clinically judged to be suffering from 'schizophrenic thought disorder' at the time of testing. However, this was not the case for a little over half the subjects. The psychiatric diagnoses which had been given at the time of testing were as follows:

'Schizophrenia', 2; 'Paranoid Schizophrenia', 3; 'Paranoid Schizophrenia with affective features', 1; 'Paranoid Schizophrenia' (later re-diagnosed merely as 'Paranoid State'), 1; 'Simple Schizophrenia', 4; 'Catatonic Schizophrenia', 3; 'Hebephrenic Schizophrenia', 2; 'Acute Relapse in chronic, hebephrenic schizophrenia', 1; 'Acute Schizophrenia following operation', 1; 'Schizophrenia with depressive features', 2 (one of these was later re-diagnosed as 'depressive disorder in a person with a marked schizoid personality').

The twenty patients in the depressive group consisted of seven males and thirteen females. It was originally hoped to make this a group of 'manic-depressive' psychotics. This aim had to be abandoned because of the shortage of such patients in the two hospitals. Very few manic patients were available, and those that were, were usually too disturbed to co-operate in such a long testing programme. Manic-depressive patients were even scarcer. Patients who had been hospitalized with both a manic and a depressive episode were too infrequently admitted to allow a group to be collected within a reasonable length of time. (The present groups took over a year to collect.) This group includes only two 'manic-depressive' cases. One patient was tested during a depressive episode, although she had had previous admissions with a manic illness. The other patient was actually tested during a hypomanic phase, although she had had episodes of depression. The remainder of the cases could be regarded as psychotic. All were selected as being reasonably typical of 'endogenous depression'. However, many of the doctors who assisted in this selection did not themselves believe that 'endogenous' and 'exogenous' depressions can be regarded as separate categories. Rather, they believed that there is a continuum from the purely endogenous to the purely exogenous, most cases falling in between. The cases selected were therefore regarded as falling distinctly towards the 'endogenous' end of the continuum. As a group they were characterized by ideas of guilt and unworthiness, retardation, depressive delusions, a history of previous depressive attacks, and the absence of any clear precipitating causes. The sample selected was not a random sample of the inpatient 'endogenous depressive' population. On average, the depressive inpatient population was considerably older than the other groups. Thus, in order to match the groups for age, the depressed patients tested were selected from the younger age range of this population. Like the schizophrenic patients, most were tested within two weeks to a month of admission. Three were inpatients at Cane Hill Hospital, and the remainder inpatients of the Maudsley Hospital. Nearly all these patients were tested prior to the onset of treatment. Unfortunately, this was not always possible. One patient had received three E.C.T. treatments prior to testing, while two others had received one E.C.T. treatment each. However, all were tested while still depressed. In spite of efforts to avoid it, six patients were being given drugs at the time of testing. In all but two cases, the dose was regarded as not large enough to affect test performance, although this judgment is subjective. Two were having 50 mg. of 'Largactil' three times daily, one was having 25 mg. of 'Drinamyl' twice daily and one was having 50 mg. of 'Pacatal' three times daily. The other two were having 150 mg. per day of 'Pacatal' when testing was commenced, but this was being increased at the rate of 50 mg. a day, so that on the final day of testing, they were having about 250 mg. daily.

The twenty neurotic patients were subdivided into a group of eight 'hysterics' and twelve 'dysthymics'. The hysteric group consisted of six females and two males. The dysthymic group consisted of seven females and five males. Of the entire neurotic group, three were inpatients of Cane Hill Hospital, the remainder inpatients of the Maudsley Hospital. Nearly all the patients in this group were having fairly long-term psychotherapy. Most were tested between two weeks and a month after admission, although a few had been in hospital considerably longer. Most were free of drugs, apart from mild night sedation, at the time of testing. However, one case of 'anxiety state' was receiving 50 mg. of 'Largactil' three times a day. One case of 'reactive depression' was receiving small doses of 'Equanil' during the day, and a second similar case, small doses of Sodium Amytal during the day. The psychiatric diagnoses given to those classified as 'dysthymic' were as follows: 'Reactive Depression', 6; 'Obsessional Neurosis', 1; 'Obsessional neurosis with anxiety', 1; 'Anxiety with depressive features', 2; 'Anxiety State', 1; 'Phobic Anxiety', 1. The patients included in the 'hysteric' group were diagnosed as follows: 'Hysteria', 1; 'Hysterical reaction with multiple somatic symptoms', 1; 'Hysterical illness in hysterical personality', 1; 'Hysterical aphonia', 1; 'Fugue state', 2; 'Hysterical multiple personality', 1; 'Hysterical personality with psychopathic tendencies', 1.

The data concerning the variables on which the groups were matched are presented in Table 1.1.

TABLE 1.1

Control data

Variable		Normals (N = 20)	Neurotics (N = 20)	Depressives (N = 20)	Schizophrenics (N = 20)
Age	Mean	32·10	32·65	37·60	32·10
	Variance	89·47	104·68	165·74	69·26
	range	18–51	19–55	19–65	18–48
	significance	F = not significant			
Mill Hill Vocabulary IQ	Mean	102·70	103·74	101·35	101·26
	Variance	158·42	200·33	225·73	157·22
	range	78–126+	80–123	80–126+	80–126+
	significance	F = not significant			
Occupational status rating scale	Mean	4·20	4·60	4·35	4·40
	Variance	2·06	1·94	2·98	2·46
	range	1–7	1–7	1–7	1–7
	significance	F = not significant			
Years of schooling and training	Mean	10·95	10·85	10·45	10·80
	Variance	10·47	3·32	4·89	3·37
	range	8–22	9–16	8–17	9–14
	significance	F = not significant			

In spite of the attempt to match the groups exactly for age, the mean age of the depressive group is still slightly higher. However, an 'F' test suggests that overall mean differences in age are not significant.

The main aim of the matching was to ensure that the groups were similar in pre-illness general intellectual and educational levels. It is clearly not desirable to match the groups on a measure of general intelligence obtained during their illness, since thought disorder can reasonably be expected to reduce efficiency on many cognitive tests, and this would be controlling a variable which we wish to examine. There is considerable evidence (Payne, 1960) that vocabulary tests scored in the conventional way are least affected by mental illness. Thus, while this is not a perfect estimate of pre-illness 'general intelligence', it is probably the best available. As can be seen the groups are almost exactly matched on the Mill Hill Vocabulary test (Raven, 1948). The 'IQ' scores quoted are based on a simple transformation from Raven's published percentile norms. The 'IQ' units in this case are made to have a mean of 100 and a standard deviation of 16 for each age-group. The Mill Hill Vocabulary test was chosen as it provides a reliable score, being based on both a 'definitions' and a 'multiple choice' vocabulary test and is standardized on an English population.

The occupational status was assessed on Belson's (1955) seven-point rating scale. The mean occupational status for each group is very close to the British average. The range in the present group is a trifle restricted, in that there are no representatives of either of the extremes of income level. This is inevitable with such small groups, as it is clearly not possible in a representative sample of twenty cases to include individuals who occur only once in 1000 of the general population. In terms of the range of occupations sampled, and the variance on this rating scale, the groups can be regarded as reasonably representative of the British population in terms of occupational status.

The groups are well matched in terms of years of formal education. This measure includes years of university, teachers' training college, technical, or other full-time education in addition to years spent at school.

V. THE TESTS USED AND THE MEAN DIFFERENCES OBTAINED

A complete list of the measures derived from the battery of tests administered is given in Table 1.2, along with the code number used to denote each variable. These code numbers will be used in subsequent tables and diagrams. (The code numbers were used to tabulate the control data as well. Thus the first test code number is 9.)

TABLE 1.2

List of variables with code numbers

Code Number	Variable
9	Mill Hill Vocabulary IQ.
10	Brengelmann Picture Recognition test score.
11	Object Classification test number of 'A' responses.
12	Object Classification test number of 'non A' responses.
13	Total time on Object Classification test.
14	Average response time for Object Classification Test.

TABLE 1.2

List of variables with code numbers (continued)

Code Number	Variable
15	Average time per 'A' response for Object Classification Test.
16	Luchins test: Number of simple solutions after set.
17	Luchins test: Number of difficult solutions after set.
18	Luchins test: difficult solution to 'control problem' before set = 1, simple = 0.
19	Epstein test, total time taken.
20	Epstein test, Overinclusion score.
21	Epstein test, Neologism score.
22	Goldstein-Scheerer Colour-Form test 'concreteness' rating.
23	Goldstein-Scheerer Object-Sorting test 'concreteness' rating.
24	Goldstein-Scheerer Object-Sorting test 'Overinclusion' score.
25	Goldstein-Scheerer Object-Sorting test. Number of unusual sortings.
26	Waves test: Average amplitude.
27	Waves test: Average wavelength.
28	Dynamometer test: maximum grip.
29	Static ataxia: number of reversals.
30	Static ataxia: total movement.
31	Shaw test: Number of 'A' responses.
32	Shaw test: Number of 'B' responses.
33	Shaw test: Number of 'C' responses.
34	Shaw test: Number of 'D' responses.
35	Shaw test: total time.
36	Shaw test: Average response time.
37	Nufferno Level test score.
38	Sum of two longest 'incorrect' or 'abandoned' item times on Level test.
39	Nufferno Speed test A1, unstressed.
40	Nufferno Speed test B1 unstressed.
41	Nufferno Speed test A2 stressed.
42	Word Association test: number of synonyms.
43	Word Association test: repetition of stimulus word.
44	Word Association test: number of multiple responses.
45	Word Association test: number of aside remarks.
46	Word association test: repetition of previous stimulus or response words.
47	Babcock test: speed of writing 'U.S.A.'
48	Babcock test: speed of writing sentence.
49	Babcock test: speed of writing name.
	(Note: scores 47, 48 and 49 were combined, and are referred to as variable 47 in the factor analysis plots).
50	Babcock 'Substitution' test average time per line.
51	Time to draw three squares.
52	Wechsler Arithmetic sub test.
53	Wechsler Picture Arrangement sub test.

TABLE 1.2

List of variables with code numbers (continued)

Code Number	Variable
54	Disproportionality score, drawing design test.
55	Average time: drawing design test.
56	USES combined manual and finger dexterity score.
57	Proverbs test 'A' score.
58	Average number of words per proverb.
59	Average reaction time per proverb.
60	Average total time per proverb.
61	Pathway I: time required.
62	Pathway I: number of errors.
63	Pathway II: time required.
64	Pathway II: number of errors.
65	Shaw test: Average 'A + B' response time.
66	Shaw test: Total number of responses.
67	Shaw test: Average time per 'A' response.
68	Static ataxia: maximum backwards sway.

(*a*) *Tests of Intellectual Speed*

It was pointed out in the introduction that Babcock (1930, 1933, 1941) regarded intellectual slowness as a primary cause of schizophrenic thought disorder. However, previous studies have suggested that depressed patients are even slower than acute or early schizophrenics. It is possible that these two groups are slow for different reasons. Depressives may be abnormally slow at all tasks, perhaps because they are constantly distracted by their depressive thoughts, whatever they are doing. Schizophrenic slowness on the other hand might be confined to intellectual tasks, because it is a by-product of overinclusive thinking. In the present study, intellectual speed was measured by the Nufferno Speed tests (Furneaux, 1956). These tests consist of Thurstone-type letter series items, the subject being required to write down the next letter of the series. Normal people, when solving a letter series problem, consider only a limited range of possible solutions. Indeed they regard only the sequence of the letters as being relevant to the task. It is possible that schizophrenics do not exclude other aspects of the letter series items in considering possible solutions. For example, they might also take into account their associations to the letters. This in itself would not necessarily produce an incorrect answer, since solutions incorporating irrelevant aspects of the material might nevertheless be rejected if they did not fit. It would, however, produce on the average a slower solution time per item, since more information would have to be considered for each item. It should also lead to an increased variability of solution times.

In the present investigation, both the Nufferno 'Level' test, and the Nufferno speed tests were used. The Level test makes use of precisely the same type of items as do the speed tests, namely Thurstone-type letter series items. In this test, each item is presented individually on a card. The subject is encouraged to take all the time he needs. The test items do not get progressively more difficult, but are arranged in cycles of seven items each, starting with easy, and ending with difficult items. The test is arranged so that each individual is given only items around the 'threshold' of his level of failure. Items he would always solve correctly or would always fail (as determined by performance on the first cycle) are not given. Thus the experience of failure (and presumably the accompanying discouragement) is controlled for all subjects. Normally the test is untimed, although in the present study the solution time for each level item was timed secretly (by the use of a wall clock with sweep second hand behind the subject). Only the first three cycles of the Level test were used.

The raw scores obtained from this test were transformed linearly for ease of computation, as shown in Table 1.3.

TABLE 1.3

Level Test raw score transformations

Raw Score	Coded Score
0–64	0
65–129	1
130–194	2
195–259	3
260–324	4
325–389	5
390–454	6
455–519	7
520–584	8
585–649	9

The results for the five groups are presented in Table 1.4.

TABLE 1.4

Nufferno Level Test coded scores (Variable 37)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	6·05	5·08	5·13	4·15	5·05
Variance	5·42	1·72	1·55	6·34	5·94
Significance		F = 1·85, not significant			

It is interesting that, although the groups were deliberately not matched on a test of general intelligence, on the grounds that thought disorder might well reduce the scores, the Nufferno Level test does not differentiate the groups significantly.

Three forms of the Nufferno speed tests (Furneaux, 1956) were used in the present study. Test A1 consists of a group of easy items of homogenous difficulty level. It was given individually under 'unstressed' conditions. That is, the subjects were told to work at their own rate, and they were unaware that they were being timed. This was followed by test B1, which is similar to A1, but consists of a homogeneous set of more difficult items. It was also given individually under 'unstressed' conditions. The third test given, test A2, is of the same level of difficulty as test A1. However, it was given individually under 'stressed' conditions. That is to say, the subjects were told to work as quickly as possible, and were aware that they were being timed.

The scores obtained from all three speed tests are 'mean log time' scores (i.e. the average of the logarithm of the solution time for each item solved correctly, excluding the average time required merely to write down the answer). The mean log time scores obtained from these tests were transformed linearly for ease of computation as shown in Table 1.5.

Acc. No. 2178
Date ......

TABLE 1.5

Nufferno Speed Tests, transformations used

Test A1		Test B1		Test A2	
Mean log time	Coded Score	Mean log time	Coded Score	Mean log time	Coded Score
0·8000–0·9230	0	0·9900–1·1280	0	0·7630–0·8730	0
0·9231–1·0460	1	1·1281–1·2660	1	0·8731–0·9830	1
1·0461–1·1690	2	1·2661–1·4040	2	0·9831–1·0930	2
1·1691–1·2920	3	1·4041–1·5420	3	1·0931–1·2030	3
1·2921–1·4150	4	1·5421–1·6800	4	1·2031–1·3130	4
1·4151–1·5380	5	1·6801–1·8180	5	1·3131–1·4230	5
1·5381–1·6610	6	1·8181–1·9560	6	1·4231–1·5330	6
1·6611–1·7840	7	1·9561–2·0940	7	1·5331–1·6430	7
1·7841–1·9070	8	2·0941–2·2320	8	1·6431–1·7530	8
1·9071–2·0300	9	2·2321–2·3700	9	1·7531–1·8630	9

The results obtained are presented in Tables 1.6, 1.7 and 1.8.

TABLE 1.6

*Nufferno Speed Test A*1 (*individual, unstressed*) *coded scores* (*Variable* 39)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	19	18
Mean	2·30	2·25	2·38	4·21	3·89
Variance	3·06	2·02	3·12	4·51	4·93
Significance			$F = 4{\cdot}05, p < 0{\cdot}01$		

137
EYS V2

EDNL. & PSYL. RESE
LIBRARY
2178
3

S.C.E.R.T., West Bengal
Library

TABLE 1.7

*Nufferno Speed Test B*1 (*individual, unstressed*) *coded scores* (*Variable* 40)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	19	18
Mean	1·95	2·67	2·00	3·68	3·33
Variance	1·94	1·88	4·00	3·34	1·88
Significance	$F = 4{\cdot}02,\ p \overset{\sim}{<} 0{\cdot}01$				

TABLE 1.8

*Nufferno Speed Test A*2 (*individual, stressed*) *coded scores* (*Variable* 41)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	19	18
Mean	1·80	2·50	2·25	3·84	3·50
Variance	2·17	2·27	3·36	4·92	5·09
Significance	$F = 3{\cdot}58,\ p < 0{\cdot}05$				

These results are as expected. As in previous studies (Payne, 1960) it is the depressives who are slowest on all three tests, although the schizophrenics are slower than the neurotics and normals. (The various group comparisons which might have been made using '*t*' tests have not been attempted, as this is not strictly permissible statistically.)

Generally speaking, we might expect individuals who are unusually slow at problem-solving to tend to compensate for this by becoming more persistent. This might reasonably be expected of the schizophrenic group, if their slowness is the result of overinclusive thinking. With so many more data to consider in order to solve a problem, the overinclusive thinker would have to become more persistent than average, if he were to reach any solution at all. However, this need not be the

TABLE 1.9

Level Test, transformation applied to persistence scores

Score in Seconds	Transformed Score
51–79	0
80–123	1
124–191	2
192–297	3
298–462	4
463–717	5
718–1114	6
1115–1729	7
1730–2685	8
2686–4168	9

case with the depressives. If they are slow because their worries distract them, they are likely also to become distracted from the task altogether, and thus give up if they cannot get an answer in a reasonable length of time. In order to examine these expectations, a measure of persistence was obtained from the Level test. This was the sum of the two longest times (in seconds) spent during the Level test on items which were subsequently abandoned or solved incorrectly. The distributions on this measure were markedly skewed for all the groups, as is generally the case with untransformed time scores. In order to keep the variances for the groups equal as nearly as possible, and because it was intended to intercorrelate all the measures, a log transformation was applied to normalize these distributions. The transformation is shown in Table 1.9.

The results for this score are given in Table 1.10.

TABLE 1.10

Nufferno Level Test persistence scores (transformed) (Variable 38)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	3·90	3·92	3·25	3·50	4·30
Variance	3·46	4·63	2·50	4·05	5·38
Significance		F = 0·57, not significant			

As expected, the schizophrenics have the highest mean persistence scores. However, the differences are not statistically significant.

(*b*) *Tests of Motor Speed*

It has been suggested that acute schizophrenics are slow only at intellectual tasks, because they are overinclusive. This was the case in the preliminary study (Payne, Matussek and George, 1959). On the other hand, depressives were expected to be slow at all types of activity. A number of simple motor speed tests were given to test this hypothesis. The tests used, all sub-tests from the Babcock-Levy (1940) test, were as follows:

1. The speed of writing 'the United States of America'. The score was merely the number of seconds required (Variable 47).
2. The speed of writing one's own name. The score was the number of seconds required (Variable 49).
3. The speed of writing a simple sentence from dictation. The sentence used, from Babcock's battery (item 12 MC) was 'I hope to leave here very soon'.
4. The Babcock-Levy 'substitution' test (test 7 in Babcock's battery). This test merely requires the subject to write the appropriate number (from 1 to 5) in a symbol, using a key provided. The test would appear to be a function of simple perceptual and motor

speed, although some rote learning is probably also involved. Babcock's scores were not used. The score used in the present study was merely the average time in seconds required to complete each line of the test, there being five lines in all.

In order to normalize the distributions, and make the variances as homogeneous as possible, logarithmic transformations were used as shown in Table 1.11.

TABLE 1.11

Transformations applied to Babcock Motor Speed Test raw scores

Time to Write 'United States of America' (Variable 47)		Babcock Sentence (Variable 48)	
Time in Seconds	Transformed Score	Time in Seconds	Transformed Score
6–7	0	5–6	0
8–10	1	7–8	1
11–13	2	9–10	2
14–16	3	11–12	3
17–22	4	13–16	4
23–28	5	17–20	5
29–37	6	21–25	6
38–48	7	26–32	7
49–62	8	33–41	8
63–80	9	42–52	9

Speed of writing name (Variable 49)		Babcock 'Substitution' Test (Variable 50)	
Time in Seconds	Transformed Score	Average time per line	Transformed Score
3	0	9·89–11·64	0
4	1	11·65–13·71	1
5	2	13·72–16·14	2
6	3	16·15–19·00	3
7–8	4	19·01–22·39	4
9–11	5	22·40–26·36	5
12–18	6	26·37–31·05	6
19–24	7	31·06–36·56	7
25–31	8	36·57–43·95	8
32–40	9	43·96–50·70	9

The results obtained are presented in Tables 1.12, 1.13, 1.14 and 1.15. As expected, on average the depressives were the slowest of all the groups. Quite contrary to prediction, however, the schizophrenics are clearly slower than the neurotics or the controls on all four tests. These results are not consistent with those obtained in the preliminary study (Payne, Matussek and George, 1959), and are not consistent with the

TABLE 1.12

Speed of writing 'United States of America': transformed scores (Variable 47)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	7	20	20
Mean	1·35	3·00	2·14	3·00	2·95
Variance	1·29	2·00	0·48	3·37	5·21
Significance			$F = 3{\cdot}43, p < 0{\cdot}05$		

TABLE 1.13

Speed of writing sentence: transformed scores (Variable 48)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	7	20	20
Mean	2·25	3·42	3·14	3·65	3·30
Variance	1·04	0·99	0·81	2·24	3·59
Significance			$F = 2{\cdot}83, p < 0{\cdot}05$		

TABLE 1.14

Speed of writing name: transformed scores (Variable 49)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	7	20	20
Mean	2·45	4·25	4·00	4·35	4·95
Variance	2·26	2·75	1·33	2·56	3·52
Significance			$F = 6{\cdot}51, p < 0{\cdot}01$		

TABLE 1.15

Babcock 'Substitution' Test: transformed scores (Variable 50)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	2·25	2·75	2·38	4·55	3·90
Variance	1·25	2·02	0·55	4·05	4·94
Significance			$F = 6{\cdot}03, p < 0{\cdot}01$		

hypothesis that slowness in acute schizophrenia is confined only to intellectual tests. Indeed these simple motor tests produce the same results as the Nufferno speed tests. Two possibilities are suggested. Schizophrenics may suffer from two quite unrelated abnormalities, a general retardation which they share with depressives, and 'over-inclusion' which they do not share with depressives. If this is the case, intellectual slowness in schizophrenia cannot be entirely attributed to overinclusion. The other possibility is that overinclusive thinking affects the performance even of simple motor speed tests. This seems less likely. Even if it did, it should affect problem-solving more than simpler tasks. Yet there is no evidence that the two sorts of speed test are in any way different, in terms of their capacity to differentiate the groups.

(*c*) *Tests of 'Concreteness'*

(1) *The Goldstein-Scheerer Sorting Tests.* Previous studies suggest that schizophrenics are 'concrete' in their performance on these tests, if Goldstein and Scheerer's (1941) scoring criteria are used. If the tests are scored in this way, however, the result will be ambiguous. Goldstein and Scheerer do not distinguish between what might be called a 'bizarre' sorting, and a failure to sort at all. Both performances are rated as 'concrete'. If the theory of 'overinclusion' is correct, and schizophrenics perceive more aspects of the test material as relevant to the task than do normals, they will have more data on which to base possible sortings. If they base their sorting on what normal people would regard as irrelevant aspects of the material, the sorting will be judged to be bizarre, or at least unusual. However, if the schizophrenic does not distinguish between a 'bizarre' sorting and a normal, 'abstract' sorting, which type of response occurs first will be completely random. If Goldstein's method of administration is followed, and the initial solution is an unusual one (e.g. 'patterning'), the subject's performance will be rated as 'concrete'. Furthermore, the subsequent test procedure will be altered, and in many cases the subject will not be allowed to continue to sort the material so that it will not be possible to determine whether or not he would later produce spontaneously an 'abstract' solution.

It was predicted from the theory of overinclusion that, if Goldstein and Scheerer's scoring system were followed, the schizophrenics would tend to be 'concrete' on these tests, by virtue of their 'overinclusive' tendencies. (This was the case in the preliminary study.) However, these tests should not differentiate as well as tests specifically designed to measure 'overinclusive' thinking. There seemed no reason to predict that the depressives would be different from the normal and neurotic groups.

Two tests of the Goldstein-Scheerer battery were used, the Colour-Form test and the Object-Sorting test. These were administered in accordance with the instructions in the manual. Although the authors give a detailed account of the types of responses on these tests which they label 'concrete', they do not give a system of scoring them. A five-point rating scale was therefore used for each test. A score of '0' indicates a completely 'abstract' performance while a score of '4' indicates a completely 'concrete' performance. These rating scales were very carefully constructed on the basis of the criteria laid down in the manual, and each level of performance was defined as precisely as possible. For example, rating '2' on the Object-Sorting Test was defined as follows: 'This score is given for subjects who require experiment IV, that is, subjects who fail to shift in experiment III (shifting), or who can think of only one acceptable alternative method of sorting (using the "acceptable" methods listed in the manual) in experiment III. However, this score is only given if the subject then succeeds with experiment IV (coercive shifting). Success in experiment IV is defined as explaining correctly (and in "abstract" terms) the first four groupings presented by the examiner, i.e. "all metal", "all red", "all toys" and "all double occurrences".'

The results obtained with these two tests are presented in Tables 1.16 and 1.17.

TABLE 1.16

'Concreteness' ratings from the Goldstein-Scheerer Colour-Form Test (Variable 22)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	0·50	0·83	0·88	1·05	2·35
Variance	1·05	1·42	2·12	1·42	2·33
Significance			$F = 4{\cdot}62, p < 0{\cdot}01$		

TABLE 1.17

'Concreteness' ratings from the Goldstein-Scheerer Object-Sorting Test (Variable 23)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	1·75	2·25	2·00	2·10	2·80
Variance	0·62	0·75	0·29	1·25	1·12
Significance			$F = 3{\cdot}31, p < 0{\cdot}05$		

These results are as expected. The depressives mean scores are very similar to those of the neurotic and normal groups. On the other hand, the schizophrenics are more 'concrete' than the other groups on both tests. In the preliminary study (Payne, Matussek and George, 1959), only the Object-Sorting test 'concreteness' rating differentiated significantly between a neurotic and a schizophrenic group. It was thought that this was because the more heterogeneous material of the Object-Sorting test favoured the production of 'overinclusive' responses which tended to be rated as 'concrete' when the Goldstein-Scheerer scoring criteria were used. In the present study, the Colour-Form test differentiated slightly better than the Object-Sorting test. Almost certainly the high 'concreteness' ratings in the schizophrenic group on both tests are related to the production of unusual responses, and not to the inability to make abstract generalizations. The evidence that this is the case will be discussed later.

(2) *The Benjamin Proverbs Test.* It has been argued that although schizophrenics are consistently reported as being 'concrete' in their

explanations of proverbs, these findings could also be interpreted as being due to 'overinclusive' thinking. Benjamin's proverbs test was given to the present groups in order to demonstrate that, when scored in the traditional way, the schizophrenic group will be moderately 'concrete' but that this score will not differentiate as well as direct measures of 'overinclusive' thinking. Furthermore, this 'concreteness' rating should be significantly correlated with other tests of 'overinclusion'. It was not expected that the depressive group would be abnormal on this test.

Benjamin's (1946) list of fourteen proverbs was used. The instructions used in the present study were as follows: 'I am going to read a few proverbs, and what I want you to do is simply give the meaning of each one as simply as you can. Let me give you an example. "Too many cooks spoil the broth." This means that when too many people are concerned in a task, they are likely to be less efficient than a single person would be. Now I am going to read some more proverbs and I want you to explain what they mean. Also tell me when you have finished.' (The proverbs were then read one at a time and the answers recorded verbatim, no other instructions being given.)

The answer to each proverb was merely rated 'abstract' or 'concrete' according to Benjamin's general criteria. The score was the total number of abstract answers given. If, however, no answer at all was given, that particular proverb was not counted (e.g. some subjects said they had never heard of some of the proverbs, and thus gave no answer. These were excluded.) When proverbs were omitted in this way, the score obtained was pro-rated so as to be a proportion of 14 (the highest possible score). For computational purposes, this score was transformed linearly as shown in Table 1.18.

TABLE 1.18

Transformation applied to the 'Abstraction' Scores from Benjamin Proverbs Test

Number of Abstract Answers	Transformed Score
0, 1	0
2, 3	1
4, 5	2
6, 7	3
8, 9	4
10, 11	5
12, 13	6

The results obtained are presented in Table 1.19.

These results are as expected. While the depressive group is a little less abstract than the normal and neurotic groups, it is probably not

TABLE 1.19

'Abstraction' rating from Benjamin's Proverbs Test: transformed scores (*Variable* 57)

	Normals	Dysthmyics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	3·40	3·17	4·00	2·75	2·25
Variance	1·73	1·24	0·86	1·78	2·51
Significance			$F = 3{\cdot}36, p < 0{\cdot}05$		

significantly different (i.e. the F ratio would not be significant were the schizophrenic group excluded). However, the schizophrenics on average are more 'concrete' than the other groups.

Evidence that this score does not imply an inability to generalize, but is related to 'overinclusion' will be presented later.

(*d*) *Tests of 'Overinclusion'*

(1) *Measures Derived from Benjamin's Proverbs Test.* If the concepts formed by schizophrenics are abnormally extensive, it is possible that, to them, a proverb means more than it does to normal people. Thus the concept it illustrates should be more difficult to define as it should be more complicated. One measure of overinclusive thinking would be the number of words required to explain each proverb. In the administration of this test, it should be noted that no time limit was imposed. The subjects were instructed to tell the examiner when they had finished, and what they said was recorded until they gave this positive indication that they had finished. The score used was merely the average number of words per answer. In order to normalize the distributions, and keep the variances for the groups as constant as possible, a log transformation was made as shown in Table 1.20.

The results are given in Table 1.21.

TABLE 1.20

Transformation applied to average number of words per proverb

Average Number of Words	Transformed Score
3·98–5·62	0
5·63–7·94	1
7·95–11·22	2
11·23–15·84	3
15·85–22·39	4
22·40–31·62	5
31·63–44·67	6
44·68–63.10	7
63·11–89·13	8
89·14–126·00	9

TABLE 1.21

Average number of words per proverb: transformed scores (*Variable* 58)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	2·50	3·17	4·25	3·00	4·45
Variance	0·68	0·70	2·21	1·37	5·31
Significance			$F = 5{\cdot}52,\ p < 0{\cdot}01$		

As expected, the schizophrenics have the highest mean score. The depressives, on the other hand, are very close to the normal group. On this measure at least, there is no evidence that the depressives are abnormally overinclusive. This measure, while producing mean differences in the expected direction, had not differentiated significantly between schizophrenics and neurotics in the preliminary study (Payne, Matussek

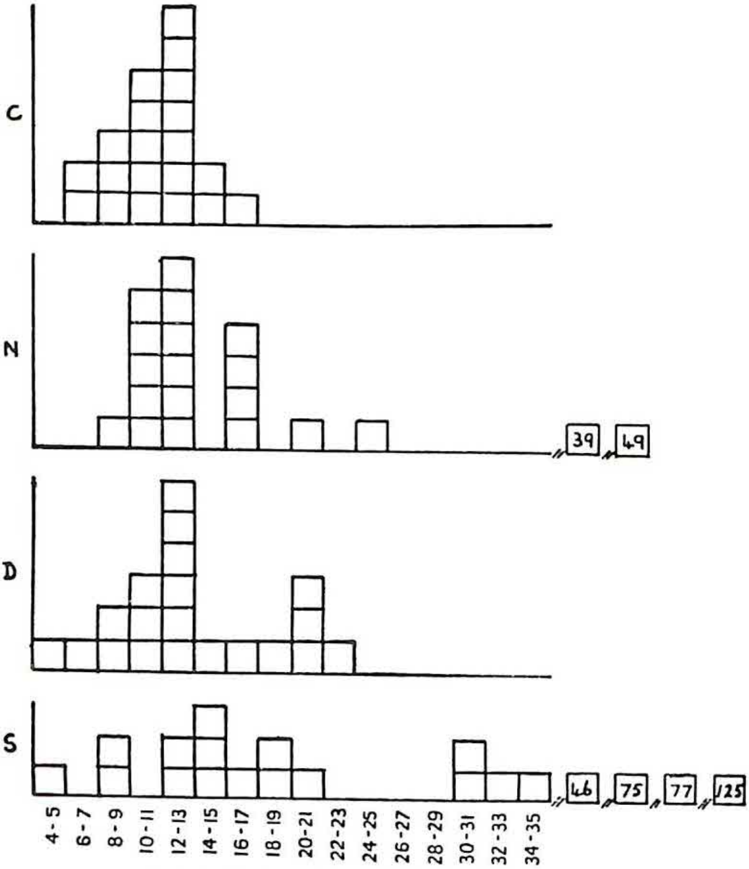


FIG. 1.1. Average No. of Words Per Proverb.

and George, 1959). It was pointed out in that study, that it is extremely difficult, when copying down patients' answers, to avoid paraphrasing them and condensing them. This tends to occur more with people who speak at length. In the present study, considerable care was taken to avoid this difficulty. If the examiner was unable to keep up with a patient's response while recording verbatim, a count of the number of words used was kept by putting a dot on the record form for each word missed. An interesting feature in these results, is that the hysterics, unlike the normals, dysthymics or depressives, use nearly as many words in their answers as do the schizophrenics. This suggests that this score may not be a pure measure of 'overinclusion'. Extraverted people, particularly those judged to have 'hysterical personalities' tend to talk more than average, and this would increase their score. Unfortunately, no other measure of extraversion was included in the present battery of tests, so this hypothesis cannot be tested on the present data. The histograms illustrating the performance of each group on this score, are presented in Fig. 1.1. Note that the scores are not transformed, but are expressed as the average number of words used.

If overinclusive subjects are explaining a more extensive and difficult concept when interpreting proverbs, in addition to requiring more words, they should need longer to prepare what they are going to say.

TABLE 1.22

Transformations applied to time scores: Proverbs Test

Mean Reaction Time		Mean Total Time	
Average Time (seconds)	Transformed Score	Average Time (seconds)	Transformed Score
2·14–3·05	0	6·38–8·51	0
3·06–4·37	1	8·52–11·35	1
4·38–6·24	2	11·36–15·14	2
6·25–8·91	3	15·15–20·18	3
8·92–12·74	4	20·19–26·92	4
12·75–18·20	5	26·93–35·89	5
18·21–26·00	6	35·90–47·86	6
26·01–37·15	7	47·87–63·83	7
37·16–53·09	8	63·84–85·11	8
53·10–75·86	9	85·12–113·50	9

That is, they should have an increased 'reaction time', if this is defined as the time (in seconds) elapsing between the presentation of the proverb by the examiner, and the first word spoken by the subject in answering. Their explanation itself should also take longer. Thus, two further measures of 'overinclusion' used, were the 'average reaction time' for each proverb and the 'average total time' (in seconds) required to give the answer. Since both scores are time measures, however, they will not be pure measures of 'overinclusion'. People who are merely slow, will tend to have abnormal scores, just as will people who are slow because they are overinclusive. It was expected that these two scores would measure 'slowness' and 'overinclusion' equally. Thus, both depressives and schizophrenics should tend to be abnormal on these two measures.

In order to normalize the distributions, and keep the variances equal for each group as far as possible, log transformations were used for these two scores, as shown in Table 1.22.

The results are presented in Tables 1.23 and 1.24.

TABLE 1.23

Mean reaction time for Proverbs Test: transformed scores (Variable 59)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	2·55	3·50	2·13	3·95	3·95
Variance	3·00	3·73	1·84	4·16	2·50
Significance			$F = 3{\cdot}08, p < 0{\cdot}05$		

TABLE 1.24

Mean total time for proverbs: transformed scores (Variable 60)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	1·80	3·17	2·50	3·35	4·74
Variance	1·75	4·15	0·86	2·87	2·76
Significance			$F = 8{\cdot}64, p < 0{\cdot}01$		

These results are more or less as expected. On the reaction time score, the depressives and schizophrenics are equally slow, the other groups all being faster. Schizophrenics are slowest in average total 'explanation' time, while depressives are less slow, although slower than the other groups. Presumably while retardation in the depressive group produces slowness on this measure, a longer answer (due to 'overinclusion') produces even greater slowness.

It is worth commenting on the scores produced by the two neurotic groups. Although the hysterics use more words, they are quicker in giving their answer than are the dysthymics. Again these differences may reflect personality characteristics, dysthymics tending to be slower, more thoughtful and more concise, hysterics tending to be faster and more verbose. These differences are compatible with what is known of the cognitive differences between these two groups (Payne, 1960). It emphasizes the possibility that the individual measures of overinclusion used in this study may measure specific factors in addition to 'overinclusion' (cf. chapter by Slater).

(2) *Measures Derived from the Goldstein Object-Sorting Test.* It has

already been argued that schizophrenics obtain their 'concreteness' rating on this test, not because they are unable to make an abstract generalization, but because they tend to produce unusual methods of sorting the objects. The number of unusual methods of sorting should provide a measure of 'overinclusion'. This measure was derived from the 'sorting' section of the test (experiments II and III in the Goldstein-Scheerer (1941) manual). An 'unusual' sorting was defined as any method of grouping the objects not listed by Goldstein and Scheerer when they describe the various 'abstract' ways in which the objects can be grouped.

The results obtained are presented in Table 1.25.

TABLE 1.25

Number of unusual sortings—Goldstein-Scheerer Object-Sorting Test (Variable 25)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	0·45	1·00	1·25	0·40	1·25
Variance	0·26	0·91	1·07	0·57	1·46
Significance			$F = 3{\cdot}64, p < 0{\cdot}01$		

As expected, the schizophrenics produce more unusual responses than do the normals. There is no evidence that the depressives are overinclusive on this measure, as their mean score is nearly identical with the normals. The hysterics unexpectedly produce as many unusual methods of sorting as do the schizophrenics. This result is specific to this test. Several other measures of unusual types of sorting were obtained, and on none of the others do the hysterics obtain extreme scores. This could be a chance finding. In any event, no explanation can be offered.

During the first stage of the administration of the Object-Sorting test (the 'Handing-Over' experiment—experiment I), the subjects are asked to select one of the objects on the table. They are then asked to hand over to the examiner all the objects which might be grouped together with this object. As each object (including the first) is handed over, it is placed out of sight by the examiner. This experiment thus requires the subject to build up a concept around the initial object. If schizophrenics tend to form overinclusive concepts, they should select significantly more objects as belonging together during this part of the experiment.

Although Goldstein and Scheerer allow the subjects to select their own objects as starting points in this stage of the experiment, Payne, Matussek and George (1959) standardized this aspect of the test to some extent. Their procedure was used in the present experiment. The 'handing-over' experiment is performed four times. On the first trial, the subject was allowed to select his own object. For the remaining three trials the examiner selected the object which was to be the 'point of departure'. The objects were always the red plate, the box of matches and the bicycle bell in that order. All the subjects were given the 'male' set of objects described in the manual, regardless of sex. The object-sorting test material used in the present investigation is shown in Plate I.

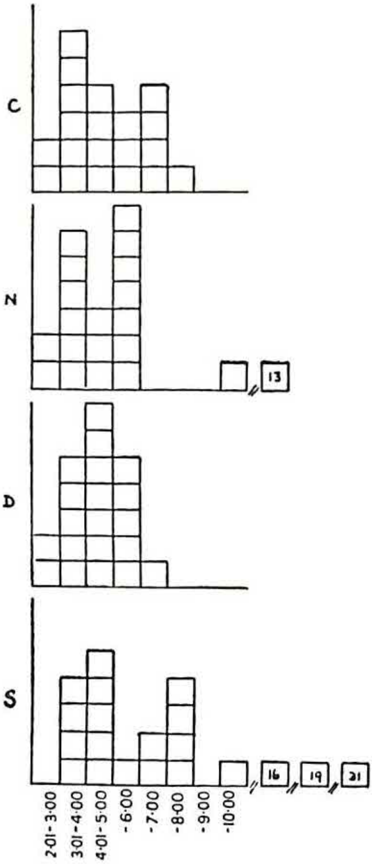


FIG. 1.2. Overinclusion Score.

The overinclusion score used was merely the average number of objects grouped together with the four objects used as 'points of departure'. Fig. 1.2 shows the histograms for the groups for the untransformed scores.

In order to normalize the distributions, and equate the group variances as far as possible, a log transformation was applied to this score as shown in Table 1.26. The results are presented in Table 1.27.

TABLE 1.26

Transformation applied to Overinclusion Scores: Goldstein-Scheerer Object-Sorting Test

Score	Transformed Score
2·23–2·91	0
2·92–3·69	1
3·70–4·92	2
4·93–6·39	3
6·40–8·31	4
8·32–10·81	5
10·82–14·06	6
14·07–18·28	7
18·29–23·76	8
23·77–31·00	9

TABLE 1.27

'Overinclusion' scores from the Goldstein-Scheerer Object-Sorting Test: transformed scores (Variable 24)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	2·10	2·33	2·88	2·05	3·75
Variance	1·78	1·51	2·12	1·00	4·62
Significance			$F = 4{\cdot}25,\ p < 0{\cdot}01$		

These results are as expected. The schizophrenics have the highest mean overinclusion score. On the other hand, the depressives obtain a mean score almost identical with that of the normal group.

(3) *The Object Classification Test*. This test was originally designed to provide an easier and more objective method of measuring 'concreteness' in sorting behaviour, as defined by Goldstein (Payne, 1954). The aim was to incorporate all the features of the three Goldstein-Scheerer sorting tests in one test.

The test consists of twelve small objects: four triangles, four squares and four circles (see Plate II). The objects vary in weight, thickness, size, material, and the hue and brightness of the colours which they are painted. There are intended to be ten different ways of grouping the objects as follows: (1) by shape (three groups, squares, circles and triangles), (2) by size (two groups, some are large and some are small), (3) by thickness (two groups, half the objects are twice as thick as the others), (4) by surface area (six groups, those of the same shape and surface size have an identical surface area), (5) by presence or absence of curved edges or corners (two groups), (6) by weight (at least three groups. This sorting must be independent of density), (7) by material (three groups, some are metal, some wood, and some a light plastic. This sorting is allowed if the subject is unable to distinguish between the wood and plastic objects, making only two groups), (8) by colour (eight groups, those of precisely the same colour are grouped together), (9) by hue (four groups, red, blue, green and yellow), (10) by saturation of colour (two groups, some colours are highly saturated and 'bright' looking, some are unsaturated and 'dull' looking).

The instructions used with the test are very simple. They are as follows: 'Do you see these objects on the table? I want you to sort them out into groups. Put them in some sort of order. Put those together which you think belong together.' (When the subject has produced one of the 10 sortings, say:) 'Good. Why did you do it that way?' (If the subject's explanation is correct, proceed with the test. Note that the explanation need not be completely 'abstract' in Goldstein's sense, just sufficiently clear to be sure the subject has really perceived that particular sorting method. If the subject cannot explain his grouping, or if he cannot produce one of the ten methods of grouping, the examiner demonstrates the 'shape' sorting, but does not subsequently credit it. Otherwise the initial sorting is credited. The final instructions are as follows:) 'Now there are other ways of sorting this material. I want you to sort the objects in as many other simple ways as you can. Arrange the objects into groups as you have just done, so that there is a single simple reason for each grouping. Each time you sort them, I want you to tell me why you have sorted them in that way.' (No further instructions are given at all, but each explanation is recorded. There is no time limit. The first time the subject spontaneously says he can do no more, he is encouraged to continue. The test is ended when he gives up a second time. All the responses are timed secretly. This was done in the present experiment by means of a wall clock with sweep second hand behind the subject. The time for each response, including the initial one, is taken from the time the subject has been told to sort the objects, or, during the main part of the test, from the time he has finished his last explanation. The time taken for him to explain his method of sorting to the examiner is not included, since presumably he is not thinking of a new method of sorting while giving this explanation. Thus each response time starts when he begins to sort, and ends when he begins his

TABLE 1.28

Object Classification Test: number of 'A' responses (Variable 11)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	4·20	4·00	3·38	3·15	2·80
Variance	2·59	2·73	2·84	4·24	5·12
Significance			F = 1·76, not significant		

explanation to the examiner. The 'total time' score is merely the sum of these response times.)

If it is true that schizophrenics are not 'concrete', but are 'overinclusive', a number of predictions can be made concerning their sorting behaviour on this test. Since schizophrenics will perceive more aspects of the material as being relevant to the task than will normal people, they will be able to elaborate many more methods of sorting. These extra methods of sorting would not be considered by normal people because they ignore the aspects of the material on which they are based (e.g. the shadows cast by the objects on the table, slight scratches on the objects, personal associations to the objects and so on). Schizophrenics should, therefore, take longer to do the test because they think of more solutions. They should not distinguish between 'normal' and 'unusual' solutions, and should, therefore, produce both types of

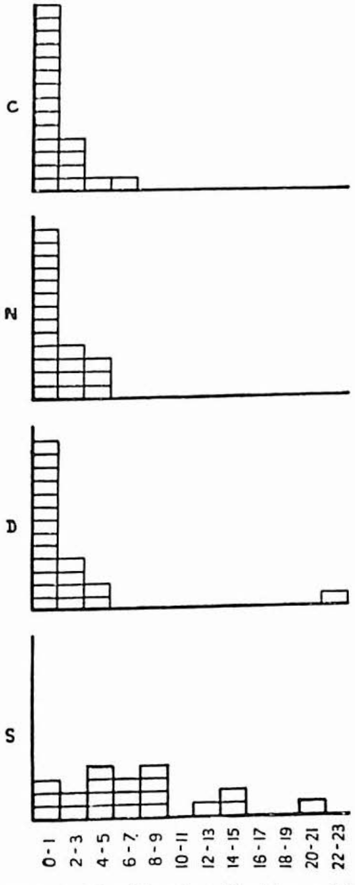


FIG. 1.3. Object Classification Test 'non-A' responses.

sorting in a random order. If there were no time limit for the test, there should be no reason why they should not produce as many 'abstract' or 'normal' methods of sorting as normal people do, although these would be randomly interspersed among the other solutions. In practice, of course, all the subjects will tend to give up the test after working on it for a period of time, so that it would not be surprising to discover that, while schizophrenics produce many more solutions, they nevertheless produce slightly fewer 'abstract' or 'usual' solutions than normal people before they give up.

A 'usual' or correct response was defined as one of the ten groupings intended in the construction of the test. It was denoted as an 'A' response. Table 1.28 gives the average number of 'A' responses obtained by the five groups.

This score could be taken as a measure of the ability to 'abstract'. As can be seen, the groups are not significantly different, although the schizophrenics have a slightly lower mean score, as expected. There is thus no evidence on this test, that schizophrenics are abnormally 'concrete'.

The 'unusual' methods of sorting this test material were called 'non-A' responses. A 'non-A' response was defined as any response not included in the list of ten 'A' responses, or a repetition of a previously given 'A' response. The 'non-A' response score should clearly be a measure of 'overinclusion'. The number of 'non-A' responses given by the groups is shown in the histograms in Fig. 1.3.

The 'non-A' response score was transformed for ease of computation as shown in Table 1.29.

TABLE 1.29

Transformation applied to number of 'non-A' responses. Object Classification Test

No. of 'non A' Responses	Transformed Score
0–2	0
3–5	1
6–8	2
9–11	3
12–14	4
15–17	5
18–20	6
21–23	7

The results for this measure are given in Table 1.30.

As expected, it is the schizophrenics who are extremely overinclusive on this measure. The depressives obtain a mean score very close to those obtained by the neurotics and the normals. Again there is no

TABLE 1.30

Object Classification Test: 'non-A' responses: transformed scores (Variable 12)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	0·20	0·33	0·13	0·60	2·05
Variance	0·27	0·24	0·24	2·46	3·63
Significance			F = 6·85, $p <$ 0·01		

evidence of abnormal overinclusion in the depressive group. It is clear from these results that, as expected, schizophrenics can obtain many more sortings than normals on this test; a normal number of 'usual' or 'abstract' sortings, as well as a large number of 'unusual' responses based on aspects of the material ignored by normal people. Space does not permit a list of all the types of unusual sortings obtained from this group. They were by and large what was predicted. For example, one man classified the objects in accordance with the type of shadows they cast on the table. As the sun moved around during the course of the test, he was able to produce a second 'shadow' classification at the end of the test, as the shadows had changed. Normal people assume that this is a two-dimensional test. That is to say, the groups must be laid out on the table. However, if a third dimension is introduced, many more groupings are possible. Some schizophrenics utilized this additional complication by forming their groups into piles of a certain order. The condition of the surface of the objects, and various types of associations to the objects were also used to produce additional methods of sorting. As predicted, the schizophrenics tended to produce usual and unusual responses in a random order, although no objective measure of this

TABLE 1.31

Transformation applied to total time, Object Classification Test

Total Time (seconds)	Transformed Score
31–48	0
49–75	1
76–115	2
116–178	3
179–275	4
276–424	5
425–654	6
655–1009	7
1010–1556	8
1557–2399	9

has been used. When unusual responses were given by the other groups, it tended to be at the end of the test, just before they gave up.

Several time scores were derived from this test. Overinclusive subjects, because they produce more classifications and have more 'data' to think about, should spend a longer total time on the test. However, the 'total time' score will not be a pure measure of overinclusion. Subjects who are merely slow will also have high scores. It was therefore expected that this score would measure both overinclusion and general retardation. Depressives as well as schizophrenics should tend to obtain extreme scores. In order to normalize the distributions, and keep the group variances relatively homogeneous, the log transformation shown in Table 1.31 was applied to this score.

The results obtained are presented in Table 1.32.

TABLE 1.32

Object Classification Test Total Time: transformed scores (*Variable* 13)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	3·95	4·42	3·50	4·45	5·95
Variance	1·84	1·72	2·57	3·84	3·00
Significance			F = 5·15, $p < 0{\cdot}01$		

The schizophrenics have the most extreme scores on this measure. As expected, depressives are also slower than normals, although the difference is small. Presumably the number of responses given is the main determinant of this score.

TABLE 1.33

Transformation applied to average time per 'A' Response—Object Classification Test

Average Time per 'A' Response (seconds)	Transformed Score
11–15	0
16–23	1
24–37	2
38–63	3
64–127	4
128–296	5
297–1000	6
1001–8000	7
8000 +	8

The second time score derived was the 'Average time per "A" response'. This was merely the total time taken, divided by the number of 'A'

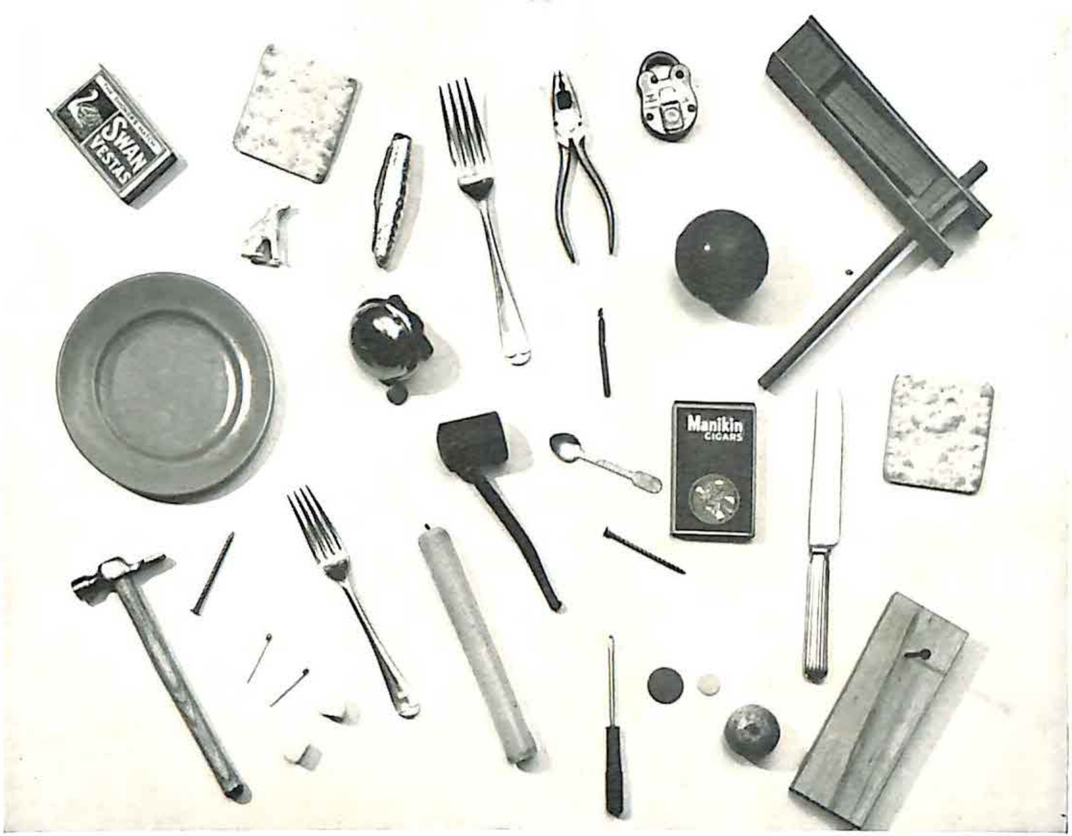


PLATE I. The Object-Sorting Test material.

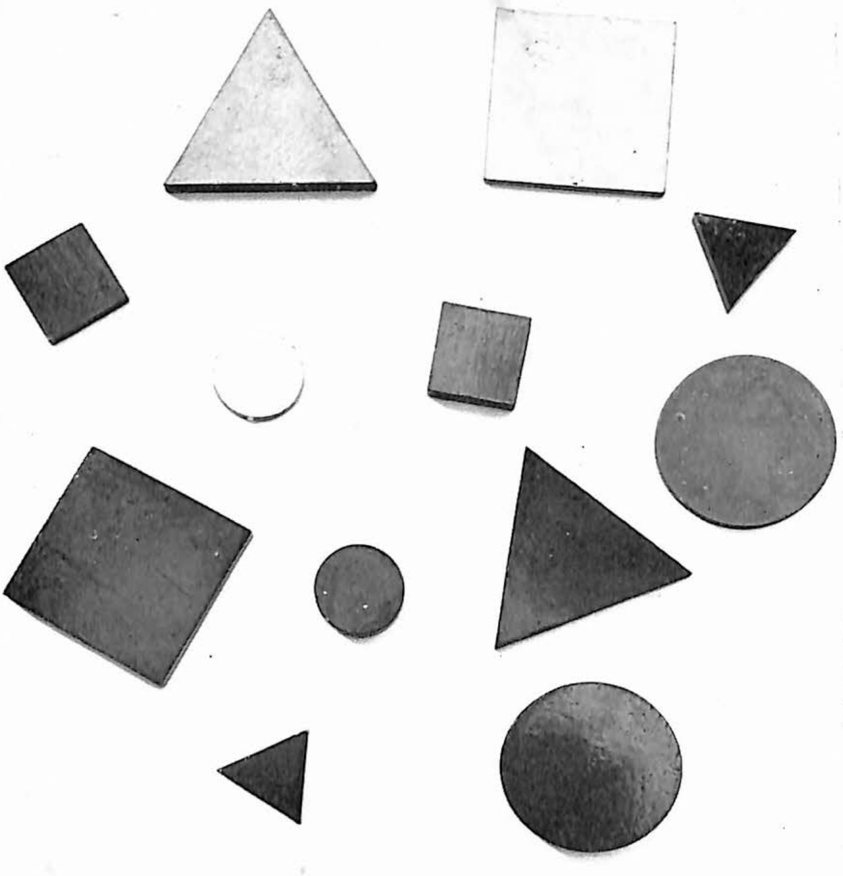


PLATE II. The Object-Classification Test material.

responses given. Individuals with a high proportion of 'non-A' responses would tend to have high scores, so that this score should be a function of 'overinclusion'. However, retarded subjects should also tend to have high scores merely because they are slow. In order to normalize the distributions, and roughly equate the group variances, the log transformation shown in Table 1.33 was applied.

The results obtained for this score are presented in Table 1.34.

TABLE 1.34

Object Classification Test average time per 'A' response: Transformed scores (Variable 15)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	3·15	3·50	3·13	4·25	5·35
Variance	1·61	1·00	1·27	3·57	3·29
Significance			$F = 6{\cdot}35, p < 0{\cdot}01$		

The results are as expected. Schizophrenics have the largest mean score, presumably because of their large number of 'non-A' responses. Depressives have a larger score than the other groups, presumably because they are slow.

The third time score derived was the average response time. This score was based on all types of sorting, and was merely the rate at which responses were produced. If schizophrenics consider more data than do normals, the test for them should be more difficult in the sense that it is more complex. For each sorting method perceived, more aspects of the material would have to be considered, both in isolating the principle

TABLE 1.35

Transformation applied to average response time scores, Object Classification Test

Average Response Time (seconds)	Transformed Score
13–16	0
17–21	1
22–28	2
29–37	3
38–48	4
49–63	5
64–83	6
84–109	7
110–142	8
143–186	9

and in allocating the objects to the groups. When more types of sorting are produced, it should also take more time to recall whether or not this method had already been given previously. Thus, overinclusive thinking should produce an increased average response time. This measure would, of course, also be a direct measure of retardation for those who are slow but not overinclusive. In order to normalize the distributions and roughly to equate the group variances, a log transformation was applied as shown in Table 1.35.

The results obtained are presented in Table 1.36.

TABLE 1.36

Object Classification Test average response time: Transformed scores (Variable 14)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	3·95	4·67	3·75	5·15	5·35
Variance	2·79	2·61	3·07	3·27	3·17
Significance			$F = 2{\cdot}56, p < 0{\cdot}05$		

As expected, the depressives and schizophrenics are nearly equally slow on this measure. It is interesting that again, the hysterics are slightly quicker than the dysthymics, a common finding in the literature (Payne, 1960).

(4) *The Shaw Test.* The Shaw Test is a test of concept formation very similar to the Object Classification Test, and it was included to replicate the findings on the Object Classification Test with slightly different test material. The set of blocks comprising the test material is illustrated in Plate III. The blocks were originally constructed by Mr. Harold Shaw of the University of Toronto. Howson (1948) used them in an investigation of intellectual impairment associated with brain damage. Bromley (1956) used them in a study of the effect of age on creative intellectual output, and Bromley's method of administration and scoring were used in the present study.

The material consists of four wooden blocks varying in colour, height, shape, size and weight. There are four shapes: hexagon, pentagon, square and triangle. There are four degrees of the colour grey, each block becoming progressively lighter than the dark grey of the hexagon. In addition, the blocks have other features by which they may be graded in a number of sequences. On the bottom of each block is one of four names: rat, lamb, ox and whale. The size of the printing varies from block to block. On the top surface, each block has one of four letters: A, B, C and D, and there is a hole in the centre of this surface varying in diameter and depth from block to block. Along the side of each block there is a notch varying in width, depth and position relative to the corner of the block.

The test instructions used were as follows: 'Would you pick up each of these blocks and examine them? You will see that they can be arranged in a kind of logical order from left to right, that is to say in a sequence or series.' (Pause.) 'There are a number of ways of arranging them. Can you find one way to begin the test?' (If the subject was unable to arrange the blocks in an acceptable manner, the 'ABCD' sequence was demonstrated, and the subject asked to repeat this. If the subject could not find a further sequence, height was demonstrated, and the subject requested to repeat this order. The subject was then asked to find as many other arrangements as possible. No time limit was used, the criterion for ending the test being the same as in the Object Classification test. The same timing procedure was also followed, to obtain individual response times, and from these, the total time.)

There are intended to be 15 'correct' or 'abstract' ways of ordering the test material. These were scored as 'A' responses, and consist of the following arrangements: (1) ABCD—by the letters, (2) Height, (3) Weight, (4) Colour, (5) Diameter of hole, (6) Depth of hole, (7) Number of sides, (8) Length of sides, (9) Depth of notch, (10) Length of notch, (11) Position of notch, (12) Alphabetical order of animals (from the names printed on the underside), (13) Number of letters in name of animal, (14) Size of letters in name of animal, (15) Size of the animal referred to.

Instead of scoring all the other possible responses together as 'non-A' responses, as in the case of the Object Classification test, Bromley has produced a scoring system for classifying them. This was followed in the present study. A 'B' response is usually a response involving some combination of the more fundamental variables. Bromley's complete list will not be given here, but some examples of 'B' responses are as follows: (1) gross size of notch, (2) reversals of 'A' sequences, (3) sequences based on very fine details, e.g. thickness of the ink in the lettering, (4) volume of the blocks, (5) gross number of edges, surfaces or corners, (6) importance of the animals to man.

In general 'C' responses according to Bromley's list are a group of less adequate responses. Examples from the list he gives are as follows: (1) grouping instead of serializing, (2) making anagrams with the single letters or attempting to make sentences with the words, (3) inane responses, (4) strength or docility of animals, etc. Many responses in this category involve the subject's associations to the objects.

'D' responses are less adequate still. A 'D' response is scored for each demonstration required at the beginning of the test. It also includes such things as describing or naming the blocks, attempting to fit them together, and building towers.

The same predictions were made for this test as were made for the Object Classification Test, the rationale being identical. The 'A' responses can be regarded as 'usual', 'abstract' responses. Theoretically, given unlimited time, schizophrenics should produce as many 'A' responses as normal people. However, the Shaw test differs from the Object Classification test in being more complex. Thus it provides more scope for a large number of unusual responses. If this occurred, the test could become so long that many overinclusive subjects would be likely to tire, and give up the test before they had exhausted all the possibilities it had for them. This would have the effect of reducing the number of 'A' responses they gave, if, as expected, from the theory, responses of all sorts are given in a random order by overinclusive subjects.

The 'A' response score was transformed linearly for ease of computation as shown in Table 1.37.

TABLE 1.37

Transformation applied to 'A' response scores, Shaw Test

Number of 'A' Responses	Transformed Score
0, 1	0
2, 3	1
4, 5	2
6, 7	3
8, 9	4
10, 11	5
12, 13	6
14, 15	7

The results obtained for this score are presented in Table 1.38.

TABLE 1.38

Shaw Test 'A' Responses: transformed scores (Variable 31)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	3·80	3·00	3·38	2·30	2·55
Variance	4·80	2·73	1·98	4·01	5·10
Significance			F = 1·68, not significant		

In spite of the fact that the schizophrenics might have been expected to give fewer 'A' responses, the differences between the groups are insignificant. Again, there is no evidence that they are abnormally 'concrete' in the sense of being unable to produce a normal, abstract generalization.

In general, it was predicted that schizophrenics should produce more of all the other types of responses than normals, because they are overinclusive. This prediction cannot be made for the 'B' responses however. Bromley introduced the category of 'B' responses to cover responses which are *usual* in normal people, but from a logical point of view are slightly inferior. They are not responses based on unusual aspects of the material. Thus, from the point of view of the present study, they can be regarded as identical with 'A' responses. The 'B' response score was transformed linearly for ease of computation, as shown in Table 1.39.

TABLE 1.39

Transformation applied to 'B' response score: Shaw Test

Number of 'B' Responses	Transformed Scores
0, 1	0
2, 3	1
4, 5	2
6, 7	3
8, 9	4
10, 11	5
12, 13	6

The results obtained for this score are presented in Table 1.40.

TABLE 1.40

Shaw Test 'B' responses: transformed scores (Variable 32)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	0·15	0·17	0·00	0·05	0·55
Variance	0·13	0·18	0·00	0·05	2·05
Significance		F = 1·25, not significant			

The number of 'B' responses do not differentiate the groups significantly. Again, there is no evidence that schizophrenics are unable to produce normal generalizations.

The 'C' responses in Bromley's scoring system are mainly comprised of responses based on 'irrelevant' aspects of the test material, or personal associations to the blocks. This score should, therefore, be a good

TABLE 1.41

Transformation applied to 'C' response score, Shaw Test

Number of 'C' Responses	Transformed Score
0	0
1–5	1
6–39	2
40+	3

measure of overinclusion. The results obtained with this score are shown in Fig. 1.4. in the form of histograms.

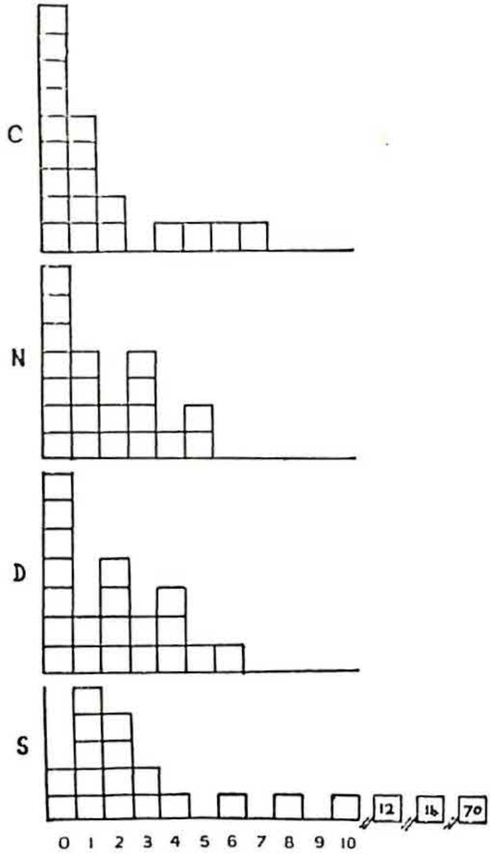


FIG. 1.4. Shaw Test. Number of C Responses.

Because of the very deviant scores obtained by some of the schizophrenics, a fourth root transformation was needed to produce approximately normal distributions on this measure. The scores were transformed as shown in Table 1.41.

The transformed scores for the five groups are given in Table 1.42.

TABLE 1.42

Shaw Test 'C' responses: transformed scores (Variable 33)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	0·60	0·58	0·75	0·70	1·25
Variance	0·47	0·27	0·21	0·33	0·51
Significance			F = 3·87, $p < 0{\cdot}01$		

As predicted, the schizophrenics have a higher mean 'C' response score than the other groups. Again there is no evidence that the depressives are overinclusive, as their mean 'C' response score is similar to those of the other groups.

The 'D' responses on the Shaw test, are not, strictly speaking, all 'responses'. This score is given when the subject needs a demonstration at the beginning of the test. It is also scored when subjects attempt to fit the blocks together, or merely build a tower. Thus 'D' responses can be scored when the subject fails to produce any method of ordering the blocks at all. In that sense, it is also a measure of 'concreteness'. The results obtained for this score are presented in Table 1.43.

TABLE 1.43

Shaw Test number of 'D' responses (Variable 34)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	0·40	0·33	0·63	1·00	1·50
Variance	0·67	0·79	1·13	2·53	4·37
Significance	F = 1·93, not significant				

The number of 'D' responses did not differentiate the groups significantly, although the schizophrenics obtained the highest mean scores, not because of their failure to respond, but because of some of their unusual responses.

It was expected from the theory of overinclusion, that schizophrenics should produce more responses of all sorts on the Shaw test, disregarding categories. The total number of responses was not normally distributed, so that the cube root transformation shown in Table 1.44 was applied.

The results for this score are given in Table 1.45.

TABLE 1.44

Transformation used with total number of responses, Shaw Test

Total Number of Responses	Transformed Score
1, 2	0
3, 4	1
5–7	2
8–12	3
13–18	4
19–27	5
28–37	6
38–49	7
50–64	8
65–82	9

TABLE 1.45

Shaw Test total number of responses: transformed scores (Variable 66)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	3·20	2·75	2·88	2·40	3·55
Variance	1·01	1·48	0·48	1·62	4 79
Significance	F = 1·75, not significant				

While the mean scores for the groups are in the expected positions, the differences overall are not statistically significant. This may be due to the length and complexity of the Shaw test. It may be that, when normal people give up the Shaw test, they have more or less exhausted the possibilities it has for them. However, the test takes so long, that overinclusive subjects tend to give up the test long before they have produced all the methods of classification they might think of. Thus they do not produce many more responses than do the normals, largely because they tire of the test. If this were the case, we should expect to find that the groups all tended to spend a similar length of time on the test. This information is available in terms of the 'total time' score. The log transformation shown in Table 1.46 was applied to this score in order to normalize the distributions.

The results are presented in Table 1.47.

TABLE 1.46

Transformation used with total time, Shaw Test

Total Time (seconds)	Transformed Score
29–46	0
47–75	1
76–121	2
122–196	3
197–317	4
318–512	5
513–827	6
828–1336	7
1337–2157	8
2158–3483	9

TABLE 1.47

Shaw Test total time: transformed scores (Variable 35)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	4·60	4·25	4·63	5·00	5·70
Variance	2·25	3·30	1·70	2·63	4·54
Significance	F = 1·69, not significant				

As expected, the groups do not differ significantly in terms of the total time spent on this test. Thus, while both the total number of responses and the total time should theoretically be partly a function of overinclusion, they may not be very valid measures when derived from the Shaw test, because of its length.

The second time score derived from the Shaw test was the average time per 'A' response. It was merely the total time, divided by the number of 'A' responses. Overinclusive individuals who have a higher proportion of 'non-A' responses, should obtain high time scores. It was thus expected that the schizophrenics would have high scores. However, a general retardation could also produce an increased 'A' response time, so that this score should not be a pure measure. Depressives should also tend to have high scores. The reciprocal cube root transformation shown in Table 1.48 was applied to this score to normalize the distributions.

TABLE 1.48

Transformation applied to average time per 'A' response, Shaw Test

Average Time (seconds)	Transformed Score
10–13	0
14–18	1
19–28	2
29–44	3
45–77	4
78–150	5
151–356	6
357–1203	7
1204–9633	8
9634+	9

The results are presented in Table 1.49.

TABLE 1.49

Shaw Test average time per 'A' response: transformed scores (*Variable* 67)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	3·60	3·75	3·75	5·00	5·35
Variance	2·78	0·57	1·07	2·32	2·03
Significance			$F = 5{\cdot}83, p < 0{\cdot}01$		

As expected, the schizophrenics have the most extreme scores, presumably because of their higher proportion of 'non-A' responses. Depressives have higher scores than the neurotics and normals presumably because of their general slowness.

As has been mentioned, on the Shaw test, 'A' and 'B' scores are nearly identical in the sense that both are 'usual' responses. Thus, the third time score examined has precisely the same rationale as the 'Average time per "A" Response'. It is the 'average time per "A" or "B" response' and is merely the total time divided by the total number of 'A' + 'B' responses for each subject. The log transformation shown in Table 1.50 was applied to this score to normalize the distributions.

TABLE 1.50

Transformation applied to average time per 'A' or 'B' response, Shaw Test

Average Time (seconds)	Transformed Score
9–16	0
17–29	1
30–49	2
50–85	3
86–146	4
147–250	5
251–429	6
430–736	7
737–1261	8
1262–2162	9

The results for this score are presented in Table 1.51.

TABLE 1.51

Shaw Test average time for 'A' or 'B' responses: transformed scores (*Variable* 65)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	2·35	2·42	2·50	3·75	4·25
Variance	2·13	0·81	0·86	2·62	3·57
Significance			F = 5·84, $p <$ 0·01		

As expected, this score produces exactly the same mean differences as did the average time per 'A' response.

The fourth time score derived from the Shaw Test was the average response time, responses of all sorts being taken into account. It was expected that on this measure, retarded people should have the most extreme scores, because of their general slowness. On the other hand, overinclusive subjects should be slower than normal, because for them the test is more complex, with many more aspects to consider each time a sorting is performed. The log transformation shown in Table 1.52 was applied to this score to normalize the distributions.

TABLE 1.52

Transformation applied to Shaw Test, average response time

Average Time (seconds)	Transformed Score
10–13	0
14–19	1
20–27	2
28–38	3
39–54	4
55–76	5
77–107	6
108–151	7
152–212	8
213–298	9

The results are presented in Table 1.53.

TABLE 1.53

Shaw Test average response time: transformed scores (Variable 36)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	3·15	3·25	3·75	4·65	4·50
Variance	1·50	1·48	1·64	2·45	3·42
Significance			$F = 3{\cdot}89, p < 0{\cdot}01$		

As expected, it is the depressives who have the most extreme scores on this measure, followed by the schizophrenics.

(5) *The Epstein Test of Overinclusive Concept Formation.* This test was developed by Epstein (1953) as a direct measure of 'overinclusion' in concept formation as defined by Cameron. The test consists of a list of fifty words printed on two pages. Following each stimulus word there are six response words, including the word 'none'. The subject is asked to underline all those response words which are an essential part of the concept denoted by the stimulus word. Epstein predicted from Cameron's theory that schizophrenics would underline more response words than normals, as their concepts would be more overinclusive. This was in fact the case in his initial study. Schizophrenics had a mean overinclusion score of 20·92, while normals had a mean score of only 12·49. Payne and Hirst (1957) tested a small group of eleven endogenous depressives and fourteen matched normals. They found the normals to have a similar overinclusion score to Epstein's group (13·79). However, the depressives were significantly more overinclusive obtaining a mean score of 37·91, considerably higher than Epstein's schizophrenic group. Payne and Hirst concluded that the Epstein test might well be measuring

a characteristic common to all psychotics. However, in the preliminary study to the present investigation (Payne, Matussek and George, 1959), the schizophrenic group had a mean overinclusion score of 56·93 on this test, the neurotics obtaining a mean of only 16·64, very close to Epstein and Payne and Hirst's normal groups. These findings are rather puzzling. On the face of it, this test should be a good measure of overinclusive thinking. Why the depressives should have obtained such high scores in Payne and Hirst's study, and why Epstein's original schizophrenic group should have obtained such low scores, it is difficult to understand, as on all the measures of overinclusion so far discussed in the present study, depressives obtained normal scores, and only schizophrenics were overinclusive. It was hoped that the present study might resolve some of these discrepancies.

The overinclusion score used in the present study is identical with that used in previous studies. It is based on those words which tended not to be underlined by normal people in Epstein's initial study, and was scored according to Epstein's procedure. In order to normalize the distributions, the log transformation shown in Table 1.54 was applied to the raw scores.

TABLE 1.54

Transformation applied to Epstein Test, overinclusion score (Variable 20)

Raw Score	Transformed Score
3–4	0
5–6	1
7–8	2
9–12	3
13–17	4
18–25	5
26–36	6
37–51	7
52–73	8
74–105	9

The results are presented in Table 1.55.

TABLE 1.55

Epstein Test overinclusion score: transformed (Variable 20)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	3·05	3·92	3·25	4·25	4·45
Variance	4·37	6·63	4·79	6·51	10·05
Significance	F = 0·97, not significant				

This is one of the few measures used in the present study which produced insignificant differences between the groups. The normals obtained a mean score similar to those found in previous studies (approximately 14 in raw score units). The neurotics obtained a very similar mean score (18) to those tested by Payne, Matussek and George. However, the depressives obtained a very much lower mean score than the group tested by Payne and Hirst (1957): approximately 22 in the present study, as contrasted with 38 in Payne and Hirst's study. The schizophrenics in the present study obtain a mean of approximately 32, closer to Epstein's original result, and markedly divergent from the mean of 57 found by Payne, Matussek and George. The present study throws very little light on the reason for these discrepancies. All the groups discussed were of similar (average) intellectual level, so that differences in intelligence cannot have produced them. The most striking feature in the present study is the extremely large variance in the schizophrenic group (see Fig. 1.5).

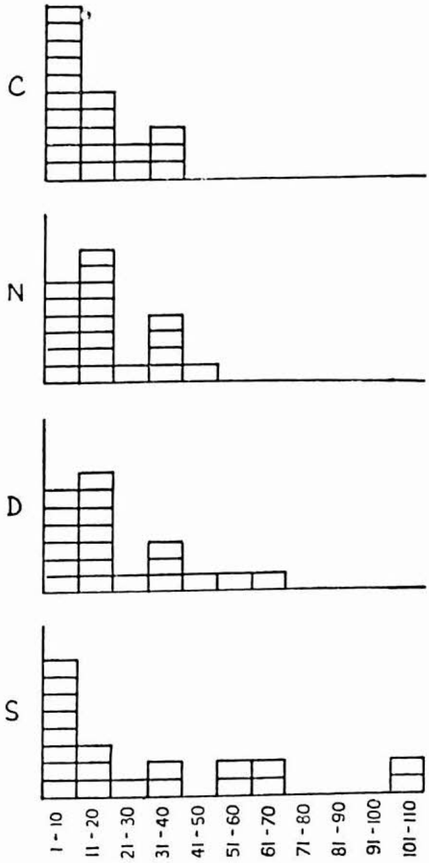


FIG. 1.5. Epstein Overinclusion Score.

This was one factor contributing to the lack of significance between the groups, and could explain why small groups of schizophrenics could vary unduly in their mean scores from sample to sample. Despite these contradictions, the means obtained in the present study are at least in the order expected from the theory of overinclusion.

The Epstein test also includes neologisms. Five of the fifty stimulus words are neologisms as are five of the response words. Epstein did not investigate the responses to these neologisms separately. In the present study a 'neologism score' was calculated, which was merely the total number of words (other than the word 'none') underlined in response to a neologism, plus the number of neologisms underlined as responses. It was thought that this score might also be a measure of overinclusion, since a high score would suggest an individual whose concepts can incorporate words which are meaningless, but which have features in common with everyday words. The neologism score was transformed linearly for ease of computation as shown in Table 1.56.

TABLE 1.56

Transformations applied to the Epstein Test neologism score (*Variable* 21)

Raw Score	Transformed Score
0–2	0
3–5	1
6–8	2
9–11	3
12–14	4

The results for this score are given in Table 1.57.

TABLE 1.57

Epstein Test neologism scores: transformed (*Variable* 21)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	0·25	0·33	0·13	0·20	0·80
Variance	0·30	0·42	0·12	0·17	1·64
Significance			F = 2·08, not significant		

Although this score produced differences in the expected direction, they are not statistically significant. Payne, Matussek and George (1959) found significant differences on this score between schizophrenics and neurotics, while Payne and Hirst (1957) found that endogenous depressives did not differ significantly from normals. Thus, in spite of the insignificant results obtained in the present study, the evidence overall still suggests that the neologism score is a measure of overinclusion, although it may not be particularly reliable.

If schizophrenics are overinclusive, they should find the Epstein Test more complicated than normal people. Their concepts should have a greater number of aspects, all of which would have to be considered. They should thus take longer to do the test. However, depressives should also spend longer at the test because of their retardation. The log transformation applied to the total time score to normalize the distributions is shown in Table 1.58.

TABLE 1.58

Transformation applied to Epstein Test total time (*Variable* 19)

Total Time (seconds)	Transformed Score
407–510	0
511–639	1
640–801	2
802–1004	3
1005–1258	4
1259–1577	5
1578–1977	6
1978–2477	7
2478–3104	8
3105–3890	9

The results for this score are given in Table 1.59.

TABLE 1.59

Epstein Test total time: transformed scores (*Variable* 19)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	11	6	17	18
Mean	2·55	2·45	1·83	4·06	5·06
Variance	1·42	2·87	3·37	4·31	5·82
Significance			$F = 6{\cdot}65, p < 0{\cdot}01$		

These results are as expected.

(6) *Luchins' Test of 'Einstellung' Rigidity*. It was hypothesized that 'overinclusion' is due to an abnormal degree of 'stimulus generalization' in schizophrenia. This should make discrimination learning more difficult. 'Set' of the 'Einstellung' type can be regarded as a special case of discrimination learning. For example, in Luchins' (1942) 'water jar' test, a typical test of 'Einstellung' rigidity, in the course of a series of problems, a specific type of complicated solution is consistently 'reinforced', simple methods not being found to work. Later, when this 'set' has been learned (discrimination learning), it is found that even when a simpler type of solution is possible, normal people continue to use the complex method due to their 'set'. If the theory is correct

schizophrenic patients should not readily be able to learn a 'set' of this type. They should not be susceptible to the development of 'Einstellung' rigidity.

In order to test this hypothesis in the present experiment use was made of a modified form of Luchins (1942) Test based on a version by Rokeach (1948).

The test items were drawn on cards each giving a representation of two (or three) containers, with a number in each to indicate its capacity. The subjects were shown Card 1, which depicted two containers, a thirty-nine gallon, and a four-gallon container. They were asked to imagine that they were beside a well, with only these two containers, and were asked how they could measure out precisely thirty-one-gallons of water. If they were unable to solve this trial problem, it was explained to them that this could be done by first filling the thirty-nine-gallon container, and then emptying out four gallons into the four-gallon container twice. The subjects were then given the 'control problem'. This problem presented three containers, a thirteen, a twenty-nine and a three-gallon container. The subjects were asked to measure out ten gallons. The 'control' problem can be solved in two ways. If the containers are labelled 'A', 'B' and 'C', it can be solved by the method A–C (the 'simple' method) or by the method B–A–2C (the 'difficult' method). The 'difficult' method is the method which was subsequently reinforced, while the 'set' was being established. The purpose of the 'control problem' was to demonstrate that all the subjects used the 'simple' method before the set was established. Following the administration of the 'control problem', another item was demonstrated, consisting of a thirty, a forty and a three-gallon container. The subjects were asked to get four gallons. It was explained that this problem was more difficult and could only be solved in one way. The 'complicated' method (B–A–2C) was then demonstrated. The subjects were then given a series of five 'set' problems, to solve by themselves. Each problem could only be solved by the 'difficult' method (B–A–2C), the purpose being to establish this set. (Rokeach used only two 'set' problems. The number was increased to five in the present study, as it was hoped that all normal subjects would be made 'rigid' by this procedure, and only

TABLE 1.60

Einstellung Test 'set' problems

A	B	C	
21	127	3	get 100
20	59	4	get 31
18	43	10	get 5
31	61	4	get 22
14	59	10	get 25

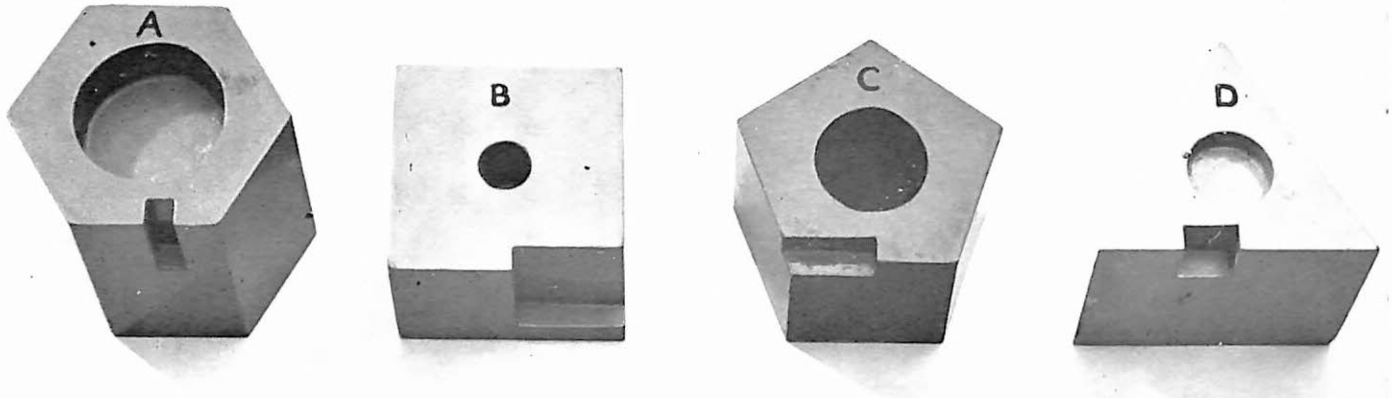
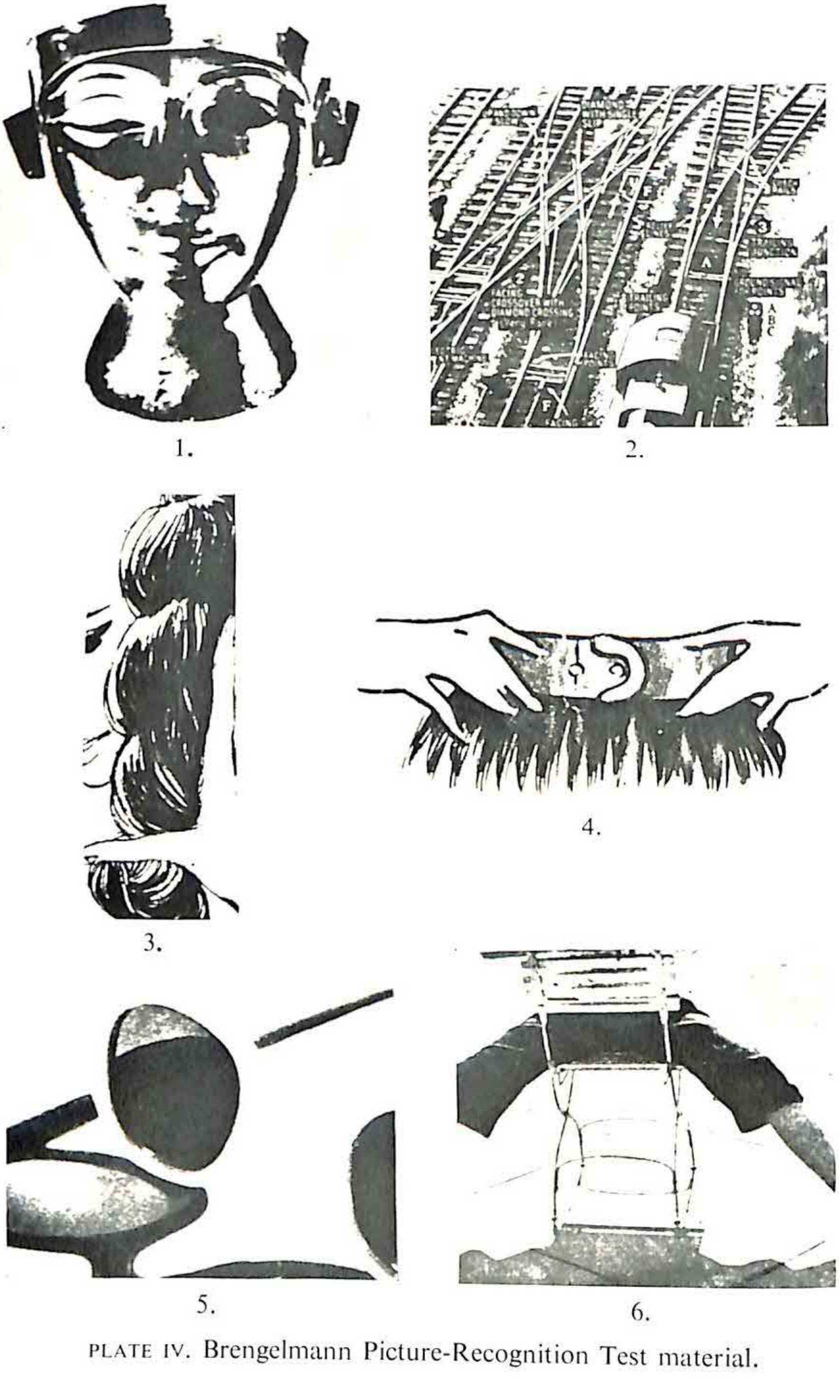


PLATE III. The Shaw Test material.



1.

2.

3.

4.

5.

6.

PLATE IV. Brengelmann Picture-Recognition Test material.

the schizophrenics would tend to be 'flexible' at the end of the test.) The five 'set' problems are shown in Table 1.60.

The subjects were then given five 'critical' problems. These followed the 'set' problems without a pause. Each of the five 'critical' problems can be solved, either by the 'simple' (A–C) method, using only two jars, or the 'difficult' (B–A–2C) method. These 'critical' problems are shown in Table 1.61.

TABLE 1.61

Einstellung Test 'critical' problems

A	B	C	
23	49	3	get 20
11	25	3	get 8
17	40	6	get 11
10	23	3	get 7
11	27	5	get 6

Following the 'set', it was expected that nearly all the subjects would use the 'difficult' method. Only those people who do not readily develop a 'set' of this sort, would use the 'simple' method. The score used was merely the number of 'simple' solutions used for the last five 'critical' problems. It was predicted that schizophrenics, being overinclusive, would get higher scores than the other groups, as they would not have developed the 'set'.

Unfortunately, a direct comparison of the groups on this measure was to some extent invalidated by the results obtained from the 'control' problem. It was assumed that, before the 'set', all the subjects would use the 'simple' method. This was not the case. Five of the twenty normals, two of the twenty neurotics, five of the sixteen depressives and four of the eighteen schizophrenics who were given the test used the 'difficult' method of solution on the 'control' item. Although these differences are not statistically significant, the groups were not exactly equal in this respect at the beginning of the test. (Note that some subjects could not be given this test because of their defective arithmetic ability.)

TABLE 1.62

Luchins 'Einstellung' Rigidity Test. Number of simple solutions given following set (Variable 16)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	16	18
Mean	1·10	1·33	1·00	1·00	1·89
Variance	3·99	2·97	3·71	3·47	3·99
Significance	F = 0·63, not significant				

The results obtained from the 'critical' items are shown in Fig. 1.6. They are also presented in Table 1.62.

The mean scores obtained from this measure are exactly in the predicted direction. All the groups except the schizophrenics obtain almost identical mean scores, the schizophrenics obtaining more simple solutions as expected. However, these differences fail to achieve the 5% level of significance, using an analysis of variance procedure. As can be seen from the histograms in Fig. 1.6, the distribution for the schizophrenic group tends to be 'U' shaped. No transformation would normalize these distributions, and the analysis of variance used is not, strictly speaking, an appropriate test of significance for this score. If we arbitrarily call 0 or 1 'simple' solutions a 'rigid' response to the test, and two or more 'simple' solutions a 'flexible' response to the test, it is possible to perform a χ^2 test on these data. If the schizophrenics (the 'overinclusive' group)

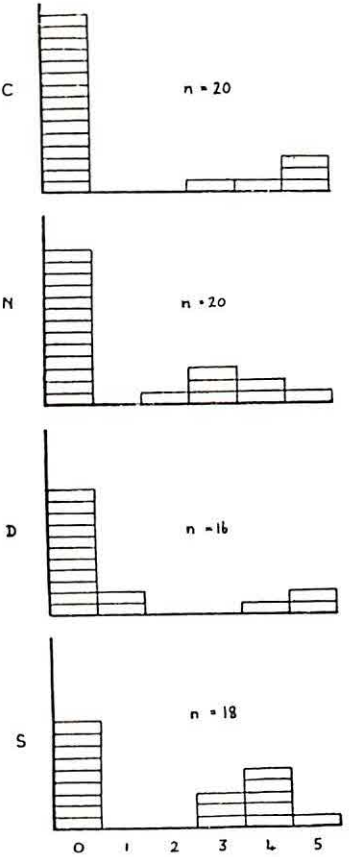


FIG. 1.6. Luchins' Modified Water Jar Test. Number of simple solutions following 'set'.

are compared with a combination of all the other groups (a 'non-overinclusive' group), we find that the schizophrenics differ significantly at the 5% level on a 'one tail' test ($\chi^2 = 3{\cdot}00$) in the direction of being more flexible. These results are regarded as supporting the theory especially in view of the fact that the Luchins Test is probably not very reliable. It is interesting that, on the above classification, half the schizophrenics are 'flexible' after the 'set', while only fifteen out of fifty-six of the remaining subjects are 'flexible' after the 'set'.

(7) *Measures Derived from the Leiter-Partington 'Pathways' Test.* It is hypothesized that 'overinclusive' subjects (schizophrenics) are unable to develop a 'set'. On the Luchins Test and on the sorting tests, this is an advantage in that 'overinclusive' subjects are not blind to possible solutions which are missed by normal people. They are less rigid and more 'original'. However, this lack of 'rigidity' or 'set' should also be a disadvantage in some situations.

In the Leiter-Partington (1950) Pathways Test, the subject is presented with a series of circles randomly drawn on a piece of paper, and randomly numbered from 1 to 10. Without giving verbal instructions, the examiner with his pencil connects the circles in numerical order. The subject is then told 'do as I did' and presented with an identical sheet of paper. If he succeeds (as nearly all subjects do), he is told 'do as I did' with a second sheet of paper (the test proper) containing many more numbered circles in a different arrangement. The authors of the test found that normal people almost never make errors, and indeed they use it as an error-free speed test. In the second part of the test, the material and procedure are identical, except that the circles have to be connected in the order 1, A, 2, B, etc., numbers and letters alternating.

The success of this test with normal people almost certainly depends upon the fact that when they perceive numbers being connected in order (or numbers and letters being connected), they immediately assume that the essence of the test is to connect the circles in numerical (or numerical-alphabetical) order. Presumably this strong 'numerical set' has been built up through past experience. They ignore the spatial juxtaposition of the circles on the paper, and the direction of the movements the examiner makes, judging these to be irrelevant. However, if schizophrenics do not adopt this normal 'set', and are unable to 'screen out' these irrelevant aspects, they might well take them into account, and, therefore, elaborate some much more complex hypothesis about what the examiner is doing, involving all these aspects. However, when they are given the proper test sheet, they find that it includes many more items, and that the spatial juxtaposition of the numbers and circles is now quite different. Any complex hypothesis involving both numerical arrangement and spatial arrangement would no longer produce a solution, since these aspects would now be in conflict. Indeed, anything more elaborate than a simple numerical sequence would not produce a solution. It was predicted that, if schizophrenics do form over-elaborate hypotheses for these reasons, they should find this test much more difficult than normal people. They should take longer to do it. Furthermore, the conflicts produced by an elaborate hypothesis which no longer 'fits', should lead to errors.

The subjects were given the two Leiter-Partington Pathway Tests according to the procedure described in the manual (Leiter and Partington, 1950). In addition to timing the performance on each pathway in seconds, an error score was calculated. An error was scored for each circle left unconnected, and each circle connected out of proper sequence. Occasionally in pathway II, a single error (e.g. an omission) could result in a reversed sequence throughout the rest of the pathway (e.g. B–2–C–3–D–4, etc., instead of 2–B–3–C–4–D, etc.). When this happened, only a single error was scored. If it happened a second time, however, each deviation from the correct sequence for the entire pathway was scored as an error.

The cube root transformation which was applied to the error score for Pathway I in order to normalize the distribution is shown in Table 1.63.

TABLE 1.63

Transformation applied to number of errors in Pathway I (Variable 62)

Errors	Transformed Score
0	0
1	1
2–8	2
9–27	3

The error score on Pathway II was transformed linearly for ease of computation as shown in Table 1.64.

The results for the two error scores are presented in Tables 1.65 and 1.66.

TABLE 1.64

Transformation applied to number of errors in Pathway II (Variable 64)

Errors	Transformed Score
0–3	0
4–7	1
8–11	2
12–15	3
16–19	4
20–23	5
24–27	6

TABLE 1.65

Number of errors in Pathway I: transformed scores (*Variable* 62)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	0·05	0	0	0·40	0·25
Variance	0·05	0	0	0·78	0·51
Significance		F = 1·40, not significant			

TABLE 1.66

Number of errors in Pathway II: transformed scores (*Variable* 64)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	0·35	0·33	0·38	0·95	1·40
Variance	1·40	0·79	0·55	3·10	5·13
Significance		F = 1·46, not significant			

Virtually none of the subjects made any errors at all on Pathway I, and the groups are not significantly different on this test. A similar result was obtained in the preliminary study (Payne, Matussek and George, 1959). It would appear that the task is so simple that, even if the schizophrenics do attend to irrelevant aspects of the demonstration pathway, they are easily able to perform the test proper. It is possible that, even if they had formulated the wrong hypothesis to account for the behaviour of the examiner on the demonstration item, the test is so simple that, as soon as they find themselves in 'conflict' they are able to perceive the simpler explanation.

TABLE 1.67

Transformations applied to time scores on the Pathways Test (*Variables* 61 *and* 63)

Pathway I Time (seconds)	Transformed Score	Pathway II Time (seconds)	Transformed Score
27–32	0	19–27	0
33–40	1	28–39	1
41–48	2	40–54	2
49–59	3	55–76	3
60–72	4	77–107	4
73–88	5	108–149	5
89–107	6	150-209	6
108–131	7	210-293	7
132–159	8	294-411	8
160–194	9	412–575	9

The error score for Pathway II produced mean differences in the expected direction, but these are not significant. In the preliminary study, this score had differentiated schizophrenics from neurotics at the 5% level. Again, it is possible that, were a more difficult test constructed with a similar rationale, it would differentiate the groups better.

If the pathways test is more difficult for overinclusive subjects, they should take longer to do it. However, general retardation should also lead to slowness on this measure. The time scores were thus not regarded as pure measures of overinclusion. In order to normalize the distributions, the log transformations shown in Table 1.67 were applied.

The results are presented in Tables 1.68 and 1.69.

TABLE 1.68

Time taken on Pathway I: transformed scores (Variable 61)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	3·55	4·45	4·13	4·80	4·68
Variance	2·68	0·67	1·27	4·17	6·12
Significance		F = 1·40, not significant			

TABLE 1.69

Time taken on Pathway II: transformed scores (Variable 63)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	3·85	4·33	4·25	5·05	4·84
Variance	2·56	2·06	1·36	3·10	4·66
Significance		F = 1·42, not significant			

These scores produce mean differences more or less as predicted. The depressives are slowest on both tests. On neither measure are the differences significant in the present study, although they significantly differentiated neurotics from schizophrenics in the preliminary study, (Payne, Matussek and George, 1959).

(8) *The Drawing Design Test*. The 'drawing design' test was developed by Beech (1956) for an investigation with brain damaged patients. It consists of a series of nine cards containing designs of the type used in a typical 'block design' test. The designs are placed on the table (which is covered with a standard background material) in front of the subject, eighteen inches from the edge of the table. He is asked to copy them in pencil in a small book, pinned at the edge of the table immediately in front of him. A fresh page is used for each fresh design. In order to investigate the accuracy of these reproductions, Beech developed a 'disproportionality' score. This score consists of the average of a number of measurements taken from each drawing, which measure the extent to which the proportions are incorrect. Beech found, testing

a group of normals, neurotics, depressives, brain damaged patients and schizophrenics, that although they worked slowly, the schizophrenics produced significantly more accurate drawings than the other groups.

It was thought that this measure could conceivably be related to overinclusion. When normal people are asked to copy a drawing in this situation, they may have the 'set' that what the examiner requires is only a reasonable facsimile. They therefore spend little time, and reproduce merely the salient features of the drawing. They disregard the detailed proportions as not being relevant in this context. Schizophrenics, on the other hand, being overinclusive, may not have this 'set'. When asked to copy the drawings, they may take this instruction quite literally, to include reproducing not only the salient features, but also all the details, as accurately as possible. Thus their work is slower but more accurate.

The disproportionality score was calculated according to Beech's (1956) method. The raw scores were transformed linearly for ease of computation as shown in Table 1.70.

TABLE 1.70

Transformation used for Drawing Design Test disproportionality score (Variable 54)

Score	Transformed Score
1·09–1·16	0
1·17–1·23	1
1·24–1·30	2
1·31–1·37	3
1·38–1·44	4
1·45–1·51	5
1·52–1·58	6
1·59–1·65	7
1·66–1·72	8
1·73–1·80	9

The results are presented in Table 1.71.

TABLE 1.71

Drawing Design Test disproportionality: transformed scores (Variable 54)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	3·10	2·67	3·25	3·40	2·50
Variance	1·88	2·24	4·79	3·73	6·05
Significance			F = 0·68, not significant		

Although schizophrenics have the lowest mean score (i.e. are most accurate), the differences are not significant. Since it has already been shown that the schizophrenics are overinclusive on a large number of tests, it appears doubtful whether unusual accuracy in this drawing test is associated with 'overinclusion', as expected. Clearly many other factors must influence performance on a test of this sort.

The length of time spent on each drawing was also measured. If schizophrenics are indeed more careful in observing details, they should be unusually slow. Depressives also should be slow, however, if they are generally retarded. The average time spent on each drawing was transformed logarithmically in order to normalize the distributions as shown in Table 1.72.

TABLE 1.72

Transformation used for average time, Drawing Design Test (Variable 55)

Time (seconds)	Transformed Score
3·98–5·21	0
5·22–6·82	1
6·83–8·93	2
8·94–11·70	3
11·71–15·31	4
15·32–20·05	5
20·06–26·24	6
26·25–34·36	7
34·37–44·98	8
44·99–58·84	9

The results are presented in Table 1.73.

TABLE 1.73

Drawing Design Test average time: transformed scores (Variable 55)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	3·30	4·50	4·75	5·30	5·55
Variance	3·06	1·73	3·93	2·33	3·42
Significance			F = 5·38, $p <$ 0·01		

In spite of the fact that the schizophrenics are not significantly more accurate, they are significantly slower on this test. The depressives, as expected, also tend to be slow. This result is very similar to those obtained from the Babcock motor tests, on which the schizophrenics were slow, even though their slowness could not reasonably be attributed to 'overinclusive' thinking. Again this result suggests that schizophrenics might be generally retarded quite apart from their abnormal overinclusion.

(9) *Inductive vs. Deductive Reasoning Tests.* It was thought likely that overinclusive thinking would affect performance on some standard cognitive tests. It would be most likely to affect tests which require the subject to perform inductive reasoning, since the categories induced would tend to be overinclusive, and occasionally distorted for this reason. It was thought on *a priori* grounds that a test likely to be so affected would be the Wechsler 'Picture Arrangement' Test. In this test, a series of pictures are placed in front of the subject in random order. The subject must first perform inductive reasoning, in that he must formulate some story which could account for all the pictures, before he can arrange them. It was thought likely that 'overinclusive' subjects would make up over-elaborate and rather unusual stories, and thus tend to obtain poor scores.

It was thought that a test unlikely to be affected by 'overinclusive' thinking, would be one requiring the subject to perform deductive reasoning, manipulating concepts which have been 'over-learned' in the past, and are thus least likely to be distorted. Numerical concepts would appear to fulfil this condition. One test of this sort would appear to be the Wechsler 'Arithmetic' sub-test.

Both these tests were administered. It was predicted that the schizophrenics would not be differentiated from the other groups by the Arithmetic test, but would tend to get low scores on the 'Picture Arrangement' test. The tests were administered and scored in accordance with Wechsler's (1944) procedure. The 'weighted sub-test' scores were transformed linearly for ease of computation as shown in Table 1.74.

TABLE 1.74

Transformations applied to Wechsler Arithmetic and Picture Arrangement scores (Variables 52 *and* 53)

Arithmetic Weighted Score	Transformed Score	'Picture Arrangement' Weighted Score	Transformed Score
1–2	0	4–5	0
3–4	1	6–7	1
5–6	2	8–9	2
7–8	3	10–11	3
9–10	4	12–13	4
11–12	5	14–15	5
13–14	6	16–17	6
15–16	7	18–19	7
17–18	8		

The results obtained with these two tests are presented in Tables 1.75 and 1.76.

TABLE 1.75

Wechsler Bellevue Arithmetic Test: transformed scores (Variable 52)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	19
Mean	5·00	4·50	5·13	3·60	4·37
Variance	5·47	5·36	1·84	3·62	3·25
Significance	F = 1·48, not significant				

TABLE 1.76

Wechsler Bellevue Picture Arrangement Test: transformed scores (Variable 53)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	3·10	2·50	3·13	2·15	2·25
Variance	2·31	1·91	2·98	2·03	1·46
Significance	F = 1·69, not significant				

The differences on neither test were significant, and these predictions cannot be said to have been borne out.

(*e*) *Tests of 'Psychoticism'*

In order to investigate the relationship between the tests of thought disorder so far discussed, and 'psychoticism', a number of tests which are thought to measure this factor were included. The tests selected were those which differentiated best between a mixed psychotic group and a neurotic group in the studies reported by Eysenck, Granger and Brengelmann (1957) and Sybil B. G. Eysenck (1955), and were regarded by the authors as good tests of 'psychoticism'.

One of the best tests of 'psychoticism' (Eysenck, Granger and Brengelmann, 1957; Eysenck, 1953c) is the Nufferno Speed Test (A1), which has already been discussed. On inspection, a large number of the 'psychoticism' tests could be classified as speed tests, since they measure intellectual, perceptual or motor speed. Others are more difficult to classify. The tests selected are described in the following section.

Speed Measures of 'Psychoticism'

(1) *The Brengelmann 'Picture Recognition Test'*. This test was developed by Brengelmann (Eysenck, Granger and Brengelmann, 1957). In its original form, it consisted of a group of three-dimensional objects viewed monocularly through a camera shutter. The purpose of the test was to discover at what shutter speed the subjects were able to recognize the objects. In the study reported by Eysenck, Granger and Brengelmann (1957), psychotics needed significantly longer shutter speeds than neurotics or normals, before they could identify the objects. The version of the test used in the present study uses pictures, rather than real objects.

This form of the test was developed by Brengelmann (1953). Pinillos and Brengelmann (1953) and Brengelmann (1958), obtained results similar to those found with the three-dimensional form of the test.

The test used in the present study consists of six pictures (see Plate IV) which were presented to the subjects tachistoscopically in a viewing box one at a time. The internal measurements of the viewing box were 1 ft. 1 in. × 8½ in. × 8½ in. The stimulus card was slid in at the back of the box at a fixed distance of 1 ft. 2½ in. from the eye position (the photographic shutter of a 3·5 epsilon aperture). The inside of the box was painted matt off white and two 100-watt bulbs were attached inside the front of the box above and below the aperture.

The subjects were told: 'I am going to place a picture in this box. I want you to look at the picture through this shutter, and tell me what it is. You can operate the shutter here yourself by this cable. Go ahead when you are ready and tell me briefly what you see—for example you might see an aeroplane, a bowl of flowers, or perhaps a man with his dog. You will get a number of tries, but if you are not sure, you can have a guess.' For pictures 3 and 6, the subject was told the number of objects that were to be identified on the card. After each response the subject was told whether he was correct or not.

The acceptable answers for the pictures were (1) face, negro head or head; (2) railway junction or railway lines; (3) face, wool and fingers or hand—all three were required; (4) hands; (5) spectacles or sunglasses; (6) person on chair—both objects were required.

The subjects were given a standard series of eleven exposures for each card, the exposure times getting progressively longer. The series of exposures was continued for each card until the subject obtained the correct answer. The exposure times were as shown in Table 1.77.

TABLE 1.77

Brengelmann Picture Recognition Test exposure times

Trial	Exposure Time
1	$\frac{1}{100}$ second
2	$\frac{1}{100}$ second
3	$\frac{1}{5}$ second
4	$\frac{1}{5}$ second
5	$\frac{1}{2}$ second
6	$\frac{1}{2}$ second
7	1 second
8	1 second
9	3 seconds
10	3 seconds
11	30 seconds

The score for each card was merely the number of exposures necessary to identify the picture. If the subject failed to identify the picture even with the thirty-second exposure (trial 11), a score of 12 was given. The total score for the test was the sum of the scores for all six pictures.

For ease of computation, this score was transformed as shown in Table 1.78.

TABLE 1.78

Transformation used for Brengelmann Picture Recognition Test scores (*Variable* 10)

Sum of Exposures	Transformed Score
21–25	0
26–30	1
31–35	2
36–40	3
41–45	4
46–50	5
51–55	6
56–60	7
61–65	8

The results obtained for this score are presented in Table 1.79.

TABLE 1.79

Brengelmann Picture Recognition Test: transformed scores (*Variable* 10)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	2·00	2·08	3·13	4·15	3·40
Variance	1·16	1·54	1·27	1·61	4·25
Significance			$F = 7{\cdot}06, p < 0{\cdot}01$		

The results from this test are very similar to those obtained with the intellectual (Nufferno tests) and motor (Babcock tests) speed tests. The depressives are slowest, followed by the schizophrenics. Again, this is consistent with the hypothesis that depressives are generally retarded. However, as with the results obtained from the motor tests, it is difficult to see how abnormal 'overinclusion' could produce so much slowness on this test in the schizophrenic group.

(2) *The 'Three Squares' Test*. This test is extremely simple. The subjects are given a plain sheet of paper, and merely instructed to draw three squares with a pencil. The score is merely the time required. This would appear to be a motor speed test, very similar to many of Babcock and Levy's (1940) simple motor speed tests. Eysenck (1952a) and Sybil Eysenck (1955) found it to be a good test of 'psychoticism'.

In order to normalize the distributions, the scores were transformed logarithmically as shown in Table 1.80.

TABLE 1.80

Transformation applied to 'Three Squares' test scores (Variable 51)

Time (seconds)	Transformed Score
3–4	0
5–6	1
7–10	2
11–16	3
17–25	4
26–38	5
39–59	6
60–90	7
91–139	8
140–213	9

The results obtained are presented in Table 1.81.

TABLE 1.81

Time to Draw Three Squares: transformed scores (Variable 51)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	7	19	20
Mean	1·10	2·25	2·43	2·84	3·25
Variance	0·73	1·48	0·95	1·36	2·83
Significance			$F = 8{\cdot}37, p < 0{\cdot}01$		

The results on this test are similar to those obtained with the other motor speed tests, excepting that it is the schizophrenics rather than the depressives who are slowest. This result is similar to that obtained with the drawing design test time scores. It is conceivable that the schizophrenics tend to be over-meticulous.

(3) *The USES Manual and Finger Dexterity Tests.* These tests are both from the United States Employment Service (1955) General Aptitude Test Battery. They were used as a measure of 'Psychoticism' by Sybil Eysenck (1955) who found that a combined score differentiated a psychotic group from a normal and neurotic group.

The Manual Dexterity test consists of a large board, in which large wooden pegs are inserted in holes. The task is in two parts. In the first section, the subject is required to move the pegs two at a time from one part of the board to another, using both hands. In the second part the subject is required to lift the pegs out of the holes with one hand, turn them over, and put them back again. The score is the number of pegs moved within the time limit for each section.

The Finger Dexterity test consists of a small board, containing rows of holes. In the holes on the upper part of the board are small rivets. On a vertical rod at the side of the board are a pile of washers. The subject is first required to move the rivets from the upper part of the board one at a time, placing a washer on each rivet, and then to insert the assembled rivet and washer into the corresponding hole on the lower part of the board. In the second part of the test, the subject is required to disassemble the washers and rivets one at a time, placing the washers back on the rod, and moving the rivets back to the top part of the board. The score is the number of rivets moved in each section within the time limit.

The tests were administered according to the instructions laid down in the manual. Following Sybil Eysenck (1955), a combined score was obtained from these tests. It was the sum of the scores obtained on the manual dexterity task (over three trials for each operation), plus the scores for assembling and disassembling on the Finger Dexterity test. That is, it is the sum of all the "raw" scores for both tests.

The scores were transformed linearly for ease of computation as shown in Table 1.82.

TABLE 1.82

Transformation applied to Combined Manual and Finger Dexterity scores (Variable 56)

Score	Transformed Score
115–129	0
130–143	1
144–157	2
158–171	3
172–185	4
186–199	5
200–213	6
214–227	7
228–241	8
242–255	9

The results are presented in Table 1.83.

TABLE 1.83

USES *Manual and Finger Dexterity. Combined transformed scores (Variable 56)*

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	7·15	6·25	6·38	4·75	5·85
Variance	1·61	1·66	1·12	4·30	4·56
Significance			F = 5·00, $p <$ 0·01		

These results are similar to those obtained with the other motor speed tests. The depressives are slowest, followed by the schizophrenics.

Miscellaneous Measures of 'Psychoticism'

(1) *The 'Waves' Test*. This test of 'psychoticism' was used by Eysenck (1952a) and Sybil Eysenck (1955), who found that it differentiated a psychotic from a normal and a neurotic group. This test might be described as a measure of expressive movement. It is difficult to see why this test should be related to any of the other measures used in the present battery.

The subjects were given a sheet of squared paper with a 'V' marked plainly in each corner. They were instructed to trace over the first 'V', keeping the hand and arm holding the pencil off the paper. They were then instructed to close their eyes and draw six more 'V's, still keeping their arm and hand off the paper. The subjects were instructed to keep their eyes shut until the examiner had covered up the tracings. With the paper still in the same position, the subjects were instructed to continue with the next 'V', tracing over it with their eyes open, as before, and then tracing six more 'V's with their eyes closed. This was repeated for each of the four 'V's.

The scores used were both measures of the size of the 'waves' drawn. They were (1) the 'average wavelength', which was the sum of the wavelengths (in millimetres) for each set, divided by four, and (2) the 'average amplitude', which was the sum of the first and last amplitude in each set of waves, divided by eight (again measured in millimetres). Previous studies had found that psychotics produced much larger waves, as measured by both these scores, than did normals or neurotics.

The raw scores were transformed linearly as shown in Table 1.84 for ease of computation.

TABLE 1.84

Transformations applied to 'Waves Test' raw scores (Variables 26 and 27)

Average Amplitude	Transformed Score	Average Wavelength	Transformed Score
16·01–18·00	0	40·01–51·00	0
18·01–20·00	1	51·01–62·00	1
20·01–22·00	2	62·01–73·00	2
22·01–24·00	3	73·01–84·00	3
24·01–26·00	4	84·01–95·00	4
26·01–28·00	5	95·01–106·00	5
28·01–30·00	6	106·01–117·00	6
30·01–32·00	7	117·01–128·00	7
32·01–34·00	8	128·01–139·00	8
		139·01–150·00	9

The results for these two scores are presented in Tables 1.85 and 1.86.

This test produced curiously inconsistent results. Only the 'average amplitude' differentiated the groups significantly, and on this measure the depressives and the dysthymics obtained the largest scores. Since the dysthymic group contains a high proportion of cases of 'reactive depression', it might be argued that individuals in a depressive mood, tend to draw large waves. However, this order is not repeated for the 'average wavelength' score. The correlation between these two scores was only 0·46, surprisingly low when it is considered that both measures are derived from the same set of waves. It is possible, of course, that neither measure is very reliable. Whatever the explanation of these results, it is clear that, so far as the present study is concerned, there is no evidence that either score can be regarded as a measure of 'psychoticism'.

TABLE 1.85

Waves Test 'Average Amplitude': transformed scores (Variable 26)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	1·70	3·50	2·63	3·95	2·55
Variance	1·06	2·27	3·70	4·12	4·47
Significance			F = 4·63, $p < 0·01$		

TABLE 1.86

Waves Test 'Average Wavelength': transformed scores (Variable 27)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	3·00	3·92	5·13	4·60	3·45
Variance	3·79	1·90	8·27	3·41	3·95
Significance			F = 2·46, not significant		

(2) *Strength of Grip.* Sybil Eysenck (1955) found that strength of handgrip, as measured by a hand dynamometer, differentiated between a normal, a neurotic, and a psychotic group. Normals were strongest, neurotics less strong, and psychotics weakest. This test was, therefore, included in the present study as another suggested measure of 'psychoticism'.

The subjects were given a small spring steel hand dynamometer, and asked to squeeze it as hard as possible with their preferred hand. When the subject had no preferred hand, he was asked to squeeze it once with both hands, and the higher measurement was recorded. The dynamometer was marked off in Kgs. of pressure exerted. This score was transformed linearly for ease of computation as shown in Table 1.87.

TABLE 1.87

Transformation applied to Maximum Grip (*Variable* 28)

Score (kgm.)	Transformed Score
9–15	0
16–22	1
23–29	2
30–36	3
37–43	4
44–50	5
51–57	6
58–64	7
65–71	8
72–78	9

The results for this score are presented in Table 1.88.

TABLE 1.88

Maximum Grip: transformed scores (*Variable* 28)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	4·00	3·17	2·00	2·30	3·45
Variance	6·42	4·15	2·25	4·54	6·05
Significance	F = 2·02, not significant				

In the present study, strength of grip did not differentiate the groups significantly. Although the differences were insignificant, it is interesting that the hysterics had the lowest mean score. This group had the highest proportion of women to men: six women and two men. It seemed likely that this could account for this result. As a further check on the possible 'contaminating' effect of slightly unequal sex distributions in the groups, the analysis of variance was repeated for the two sexes separately. The F ratios obtained were 2·11 for men, and 1·54 for women, neither being significant. This failure to repeat Sybil Eysenck's (1955) findings may well be due to the fact that in the present study, the groups were matched for age. This was not the case in Sybil Eysenck's research. Her normal group had a mean age of twenty-two, her neurotic group a mean age of thirty and her psychotic group a mean age of thirty-nine. (These groups were matched for sex.) The strength of grip score placed the groups in the reverse order to their mean age, and it seems likely that the results were merely due to these age differences, i.e. the younger subjects tended to be stronger. There is thus no evidence that strength of handgrip is related to 'psychoticism'.

(3) *Static Ataxia.* Sybil Eysenck (1955) found that when she asked her subjects to stand with their feet together and their eyes closed for thirty

seconds, the psychotics swayed more than the other groups. Two measures differentiated the groups significantly, the 'number of reversals' (i.e. the number of separate to and fro movements) and the maximum extent of backwards sway. In the present investigation, the subjects were asked to take off their shoes, and stand with their feet together and their hands by their sides. They were then asked to close their eyes, and their movements were measured for a thirty-second period. This was done by means of a recording apparatus activated by a string clipped on to the subject's collar. Four measures were initially taken: (1) The total number of reversals, or changes of direction of movement; (2) the total movement—i.e. the distance between the largest backward and forward sway; (3) the maximum backward sway from the original upright position; and (4) the maximum forward sway from the original upright position. Due to a failure of one part of the recording apparatus, the fourth score (maximum forward sway) had to be abandoned.

TABLE 1.89

Transformation applied to Static Ataxia 'Number of Reversals' (Variable 29)

Reversals	Transformed Score
1–3	0
4–6	1
7–9	2
10–12	3
13–15	4
16–18	5
19–21	6
22–24	7
25–27	8

TABLE 1.90

Transformation applied to Static Ataxia 'Total Movement' score (Variable 30)

Total Movement (in half inches)	Transformed Score
1–2	0
3–4	1
5–7	2
8–10	3
11–14	4
15–18	5
19–23	6
24–29	7
30–35	8
36–42	9

One subject, a female 'hysteric', fell over backwards during the course of the test. She was arbitrarily given the maximum possible backward sway score.

The 'number of reversals' score was transformed linearly for ease of computation as shown in Table 1.89.

A square root transformation was applied to the 'total movement' score to normalize the distributions as shown in Table 1.90.

A 'hyperbolic tangent' transformation was applied to the 'maximum backward sway' score, to normalize the distributions as shown in Table 1.91.

TABLE 1.91

Transformation applied to Static Ataxia Maximum Backwards Sway score (*Variable* 68)

Backward Sway (in half inches)	Transformed Score
0	0
1	1
2	2
3	3
4	4
5	5
6–7	6
8–11	7
12–∞	8

The results for these scores are presented in Tables 1.92, 1.93, and 1.94.

TABLE 1.92

Static Ataxia, number of reversals: transformed scores (*Variable* 29)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	1·65	2·58	3·25	2·25	2·45
Variance	1·29	2·63	4·50	2·09	1·94
Significance	F = 1·98, not significant				

TABLE 1.93

Static Ataxia, total movement: transformed scores (*Variable* 30)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	1·95	3·00	3·63	2·60	3·00
Variance	1·73	2·73	7·12	3·09	2·32
Significance	F = 1·84, not significant				

There is no evidence that a measure of swaying, as such, differentiates the groups significantly. Neither the 'number of reversals' nor the 'total movement' produced significant differences. In this respect, the present results are not consistent with Sybil Eysenck's (1955) earlier findings. However, the maximum backward sway does differentiate the groups at the 5% level, and it is only the schizophrenics who have extreme scores. It is a pity that the measure of forward sway had to be abandoned. The schizophrenics are not significantly different in terms of their total movement. Thus, in order to sway further backwards than the other groups, they must have swayed less forwards. This finding is a very curious one, and it is difficult to see how it might be related to the other abnormalities investigated in the present study. It is of interest that this was the only test of 'psychoticism' which was not merely a test of motor, mental or perceptual speed, which differentiated the groups significantly in the present study.

TABLE 1.94

Static Ataxia Maximum Backward Sway: transformed scores (Variable 68)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	20	20
Mean	4·05	4·92	3·63	4·45	6·05
Variance	4·05	2·27	8·84	6·79	3·21
Significance			$F = 2{\cdot}96, p < 0{\cdot}05$		

(4) *The Word Association Test.* The word association test has been used as a diagnostic instrument for many years, and a large number of studies has been carried out with various clinical groups. It is not proposed to review this literature here. The word association test used in the present study was chosen because Sybil Eysenck (1955) regarded it as a good test of 'psychoticism'. A number of measures differentiated between her psychotic and normal and neurotic groups. She found that psychotics: (1) more frequently repeated the stimulus word; (2) more frequently gave multiple responses; (3) more frequently perseverated the same response; (4) more frequently made 'aside remarks' and (5) more frequently gave unusual associations. Several of these findings had previously been reported as characterizing schizophrenic patients.

It was thought that a number of these responses to the word association test might be associated with overinclusion. For example, the tendency to give more 'distant' or 'unusual' associations might well be due to abnormally inclusive concepts. If words mean more to overinclusive subjects, they are likely to have a wider range of associations to them. Indeed, Moran (1953) used a word association test as one measure of overinclusion, and found that the schizophrenics who gave more 'distant' associations to a set of words, tended to give the same words an 'overinclusive' definition. The tendency to give multiple responses, and aside remarks, and to repeat the stimulus word could conceivably be due to the failure to adopt the normal 'set' which prevents normal people from including such things in their response repertoire, presumably because they assume that it is not allowed.

The thirty stimulus words used in the present study were the same as those used by Sybil Eysenck (1955). They were as follows: cold, foot, butter, music, red, dream, sweet, soft, sleep, dark, quiet, hungry, anger, beautiful, man, loud, long, blue, slow, working, child, heavy, rough, memory, house, trouble, religion, hard, deep and wish. Each stimulus word was printed on a card. The instructions were as follows: 'I am going to show you some cards on which you will see a single word written. As soon as you see the word I want you to tell me the first word that comes to your mind. It does not matter what the word is, but it must be the first word you think of, and you should give it as quickly as possible.' The cards were shown one at a time. As soon as the subject responded, the next card was presented.

The following scores were obtained: (1) The number of synonyms given as responses; (2) the number of repetitions of the stimulus word; (3) the number of multiple response (more than one word); (4) the number of aside remarks given, and (5) the number of repetitions of previous stimulus or response words given.

TABLE 1.95

Transformations used for Word Association Test 'Number of Synonyms' and Number of Repetitions of Stimulus Words (Variables 42 *and* 43)

Number of Synonyms	Transformed Score	Number of Repetitions of Stimulus Word	Transformed Score
0–1	0	0–3	0
2–3	1	4–7	1
4–5	2	8–11	2
6–7	3	12–15	3
8–9	4	16–19	4
10–11	5	20–23	5
12–13	6	24–27	6
		28–31	7

The 'number of synonyms' and the 'number of repetitions of the stimulus word' were transformed linearly for ease of computation as shown in Table 1.95.

The square root transformation which was applied to the 'number of multiple responses' in order to normalize the distributions is shown in Table 1.96.

TABLE 1.96

Transformation applied to 'Number of Multiple Responses' Word Association Test (Variable 44)

Multiple Responses	Transformed Scores
0	0
1–2	1
3–6	2
7–12	3
13–20	4
21+	5

The 'number of aside remarks' and the 'number of repetitions of previous stimulus or response words' were transformed linearly for ease of computation as shown in Table 1.97.

TABLE 1.97

Transformations applied to 'Aside Remarks' and 'Repetitions of previous Stimulus or Response' Word Association Test (Variables 45 and 46)

Aside Remarks	Transformed Scores	Repetitions	Transformed Scores
0–1	0	0–1	0
2–3	1	2–3	1
4–5	2	4–5	2
6–7	3	6–7	3
8–9	4	8–9	4
10–11	5	10–11	5
12–13	6	12–13	6
14–15	7	14–15	7
16–17	8		
18–19	9		

The results for these scores are presented in Tables 1.98, 1.99, 1.100, 1.101 and 1.102.

TABLE 1.98

Word Association Test Number of Synonyms: transformed scores (Variable 42)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	19	20
Mean	1·30	1·50	0·75	0·84	1·10
Variance	1·80	1·55	0·79	1·03	2·83
Significance	F = 0·71, not significant				

TABLE 1.99

Word Association Test Number of Repetitions of Stimulus Word: transformed scores (*Variable* 43)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	19	20
Mean	0·55	1·00	2·50	2·47	2·35
Variance	1·94	4·91	10·57	7·37	10·56
Significance		F = 2·13, not significant			

TABLE 1.100

Word Association Test Number of Multiple Responses: transformed scores (*Variable* 44)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	19	20
Mean	0·65	0·67	0·75	1·37	0·75
Variance	1·40	0·79	1·93	2·13	1·78
Significance		F = 1·01, not significant			

TABLE 1.101

Word Association Test Number of Aside Remarks: transformed scores (*Variable* 45)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	19	20
Mean	0·60	0·67	0·75	1·21	1·30
Variance	0·99	0·79	2·21	2·51	4·54
Significance		F = 0·79, not significant			

TABLE 1.102

Word Associaton Test Number of Repetitions of Previous Stimulus or Response Words: transformed scores (*Variable* 46)

	Normals	Dysthymics	Hysterics	Depressives	Schizophrenics
Number of cases	20	12	8	19	20
Mean	1·35	0·75	0·75	1·32	1·15
Variance	0·77	0·57	0·50	2·67	1·61
Significance		F = 0·81, not significant			

None of these measures differentiate the groups significantly, nor are the mean positions of the groups in any consistent order. The predictions for this test are thus unverified, and Sybil Eysenck's earlier results are, for some reason, not repeated in the present study.

VI. SUMMARY OF THE RESULTS OF THE ANALYSES OF VARIANCE

(1) In the battery of tests given, there were thirteen measures of 'overinclusion' believed not to be influenced by other factors, such as speed. These thirteen measures all produced mean differences in the expected direction, in that the schizophrenics always obtained the most extreme scores. Five of these measures produced differences between the groups significant at beyond the 1% level, and one measure produced differences significant at the 5% level, while the remaining seven measures did not differentiate the groups significantly. On the whole, these findings support the hypothesis that schizophrenics are abnormally 'overinclusive'. On none of these measures did the other groups show marked differences. There was no evidence that the depressives or the two neurotic groups are abnormally overinclusive. This type of thought disorder is apparently confined to schizophrenic patients.

(2) Five measures of 'concreteness' were obtained. Three differentiated the groups significantly at the 5% level or better. These were the ratings derived from the two Goldstein Sorting Tests, and the Benjamin Proverbs test. Only the schizophrenics were abnormally 'concrete' on these measures, there being no differences between the other groups. These results were not attributed to the schizophrenics' inability to 'abstract', but to their tendency to produce unusual responses on these tests, which are arbitrarily classified as 'concrete' when the traditional method of scoring is used. Other sorting tests given suggested that, in terms of the capacity to produce 'usual' abstract generalizations, when no time limit is imposed, there are no differences at all between the groups. Neither of the psychotic groups are 'concrete'.

(3) Eleven measures of 'overinclusion' based on time scores were obtained. Schizophrenics were expected to obtain large (slow) scores because of their 'overinclusion', and depressives because of their general retardation. Nearly all these scores differentiated the groups at beyond the 1% level of significance. The depressives and the schizophrenics tended to be equally slow on these measures. The neurotics tended to obtain normal scores.

(4) Three tests of reasoning speed were included. It was predicted that schizophrenics would be slow because of their 'overinclusion' and depressives because of their retardation. The depressives were slightly slower than the schizophrenics, although both groups were considerably slower than normal. All the tests differentiated the groups at beyond the 5% level, and two of these at beyond the 1% level of significance. There was no evidence of abnormal slowness among the neurotics.

(5) Seven measures of motor, and perceptual slowness were included. It was predicted that the schizophrenics would obtain normal scores on these tests, as it was hypothesized that slowness in schizophrenia is a

by-product of overinclusive thinking, and would thus be confined to intellectual tasks. This prediction was clearly disproven. Both the depressives and the schizophrenics tended to be abnormally slow on these tests, there being very little difference between them. This finding suggests that schizophrenics as a group may suffer from two unrelated abnormalities, a general retardation which they share with depressives, and overinclusion, which is specific to schizophrenia. There was no evidence for abnormal motor or perceptual slowness in the two neurotic groups.

(6) Of the seventeen measures which have been used previously as tests of 'psychoticism', three are tests of intellectual speed (the Nufferno tests), and three are tests of motor or perceptual speed. All six speed tests differentiated the groups significantly. The depressives tended to be slightly slower than the schizophrenics, both groups being abnormally slow. The neurotics had normal scores on these tests. Of the eleven 'miscellaneous' measures of 'psychoticism' used (none of which were speed tests) only one differentiated the groups significantly at the 5% level. This was the 'maximum backward sway' on the 'Static Ataxia' test. There appeared to be no obvious relationship between this score and the other measures of thought disorder used.

VII. THE INVESTIGATION OF THE CO-VARIANCE OF THE TESTS

(*a*) *The Centroid Analysis.* Although the tests of 'overinclusion' tend to differentiate the groups as expected, it was necessary to demonstrate that they intercorrelate positively, that they thus measure a common factor, and that this factor also differentiates the groups in the way predicted from the theory of overinclusion.

In order to demonstrate this, all the groups were combined, product-moment correlations were calculated between all the measures, and the resulting matrix of intercorrelations was factor analysed. Before describing this procedure in more detail, it is necessary to discuss briefly a criticism of this method of analysis which is often made. It has already been demonstrated that most of the measures obtained produce significant mean differences between the groups. In other words, the groups cannot be regarded as coming from the same population. Since the tests are known to differentiate the groups significantly, if the groups are then combined and correlation coefficients computed, 'spuriously high' correlations will be obtained. According to this line of argument, the correct procedure would have been to intercorrelate the tests separately for each group. This would have given a more valid estimate of the 'true' correlation between the measures. This point of view was not accepted. It seems equally correct to argue that, had the correlations been computed for each group separately, the correlations obtained would have been 'spuriously low', because of the arbitrary restriction of the range of variation of the measures. A very similar problem could arise in an experiment in educational psychology. An educational psychologist might be concerned with investigating for a ten-year-old school population, the differences to be found between pupils in an 'A', a 'B' and a very retarded 'stream'. He might administer a group of cognitive tests to the pupils, and discover that the three 'streams' were very significantly different with respect to their intellectual level on most of the tests. The 'A stream' might tend to be very bright, and the 'retarded stream' very dull. Indeed, this might have been what he had predicted, since his initial hypothesis might have been that allocation to a 'stream' is largely a function of native ability. The psychologist might then wish to factor analyse his tests, in order to discover to what extent their co-variance might be attributable to a single factor. Very few people would argue that he ought not to combine all his ten-year-olds in order to investigate his test co-variances. In fact it would be a justifiable comment that, if he factor-analysed each 'stream' separately, he would be unduly restricting the range of variation, and seriously underestimating the co-variance of his tests and the importance of any general factor that might emerge.

This hypothetical example is believed to be relevant in many ways to the present investigation. Many of the tests of 'overinclusion' used in the present study were primarily designed to be measures of an abnormality of function. It was hoped that all normal people would have more or less the same score, while only abnormal people would have extreme scores. In other words, they were deliberately designed to produce no variance in a normal group. It would thus be incorrect to investigate their co-variance in a normal group only, or in any other single group in which the variance could be expected to be extremely limited. The only condition which should be fulfilled, is that the distributions on all the tests for the combined group on which the correlations are to be obtained should be normal (as should be each bi-variate distribution). The main purpose of transforming so many of the scores in the present study was to ensure that this would be the case as far as possible.

Not all the measures discussed were included in the factor analysis. In some cases, two measures were not mathematically independent of each other, and one had to be excluded. Table 1.103 lists the measures which were retained for the correlational study. As can be seen by comparing this table with Table 1.2, the following measures were omitted: Variable 13 (Total time on the Object Classification Test. This was omitted because it is a function of the other time score included), Variable 17 (Luchins Test: number of 'difficult' solutions following 'set'. This variable was initially coded by mistake. It is merely the inverse of the number of 'simple' solutions following 'set'), Variable 36 (Shaw test, average response time. This was omitted because it is a function of the other time scores included), Variable 67 (Shaw test, average time per 'A' response. This was excluded for the same reason as Variable 36). Variables 47, 48 and 49, the scores for the three Babcock Motor Speed tests (Writing name, writing sentence, writing 'U.S.A.') were combined on the assumption that they measure the same function, and that a combined score would be more reliable. The combined score was merely the sum of the transformed scores for the three tests. This variable will subsequently be referred to as Variable 47.

As can be seen from the tables so far given, on several of the tests, one or two subjects returned no score. This is almost inevitable in such a large-scale investigation with abnormal people. This was usually due to a sudden complete unwillingness to co-operate. In obtaining the correlations for these few variables, the number of cases was slightly reduced. While it is not strictly correct to include any subjects who did not produce scores on all the tests, this happened so rarely that it was judged not to have any effect on the overall results.

The present battery of tests was chosen to measure two main aspects of thought disorder, 'Overinclusion' and 'General Slowness or Retardation'. These two 'factors' should be the two major sources of co-variation within the matrix. However, a third important source of co-variation cannot be ignored. The groups were matched as closely as possible for pre-illness general intellectual level, so that no mean differences could be attributed to this factor. Nevertheless, in this battery of mainly cognitive tests, this is bound to be an important source of co-variation. No other sources of co-variation were regarded as important in the present study. Scores on the tests of 'concreteness' for example, were expected to be partly a function of 'overinclusion' (because the schizophrenics were judged 'concrete' on the basis of their overinclusive responses), but mainly a function of general intellectual level. The various tests of 'psychoticism' were thought to be measures of either 'slowness' or 'overinclusion', in so far as they measured anything at all relevant to the discrimination between the groups.

Since these three factors were the only important sources of co-variation which could be specified on theoretical grounds before the factor analysis was initiated, it was only thought justified to extract three factors from the matrix. If the theories behind the test selection were correct, it should be possible, by the appropriate rotations, to demonstrate that these three hypothesized factors could be identified clearly, and that factor scores derived from them should differentiate the groups as predicted.

The matrix of intercorrelations was factor analysed by means of Thurstone's (1947) method for extracting centroid factors. In view of the subsequent analyses performed, it would have been strictly more correct, to have performed a principal component analysis (see chapter by P. Slater). However, with such a large matrix of correlations, the first

TABLE 1.103

Factor Saturations

Variables	Factor Unrotated Centroids I	II	III	Rotated Factors Factor 'O'	Factor 'P'	Factor 'R'	Factor 'G'
9. Mill Hill Vocabulary 'IQ'	−0·36	0·46	−0·31	−0·37	0·10	−0·04	−0·54
10. Brengelmann Picture Recognition Test	0·38	0·05	0·23	0·13	0·29	0·37	0·22
11. Object Classification Test No. of 'A' Responses	−0·67	0·27	0·03	0·06	−0·25	−0·24	−0·68
12. Object Classification Test 'Non A' Responses	0·68	0·29	−0·27	−0·46	0·67	0·54	0·33
14. Object Classification Test Average Response Time	0·49	0·35	0·04	−0·16	0·59	0·57	0·12
15. Obj. Class. Test: Average Time per 'A' Response	0·75	0·16	−0·15	−0·32	0·62	0·54	0·46
16. Luchins Test: No. of Simple Solutions after 'Set'	−0·22	0·13	−0·07	−0·07	−0·05	−0·08	−0·24
18. Luchins Test: Difficult Solutions before 'Set'	0·22	0·04	0·39	0·31	0·18	0·32	0·10
19. Epstein Test: Total Time	0·58	0·38	0·10	−0·13	0·67	0·67	0·16
20. Epstein Test: Overinclusion Score	0·56	−0·42	−0·07	−0·03	0·07	0·06	0·70
21. Epstein Test: Neologism score	0·61	−0·09	−0·32	−0·37	0·35	0·23	0·54
22. Goldstein Colour-Form 'Concreteness' rating	0·71	−0·03	−0·29	−0·38	0·46	0·35	0·57
23. Goldstein Object Sorting 'Concreteness' rating	0·68	−0·11	−0·07	−0·15	0·38	0·36	0·58
24. Goldstein Object Sorting 'Overinclusion' score	0·06	0·36	−0·44	−0·54	0·31	0·13	−0·16
25. Goldstein Object Sorting unusual sortings	0·10	0·20	−0·46	−0·51	0·22	0·04	−0·02
26. Waves Test: Average Amplitude	0·41	0·00	−0·12	−0·18	0·28	0·23	0·31
27. Waves Test: Average Wavelength	0·30	−0·07	−0·16	−0·18	0·15	0·10	0·28
28. Maximum Grip: Dynamometer	−0·13	0·34	0·09	−0·00	0·16	0·17	−0·33
29. Static Ataxia: No. of Reversals	0·25	0·14	−0·21	−0·28	0·27	0·18	0·11
30. Static Ataxia: Total Movement	0·29	0·04	−0·20	−0·25	0·23	0·15	0·21
31. Shaw Test: No. of 'A' Responses	−0·72	0·47	−0·19	−0·20	−0·14	−0·23	−0·83
32. Shaw Test: No. of 'B' Responses	−0·08	0·34	−0·40	−0·47	0·20	0·04	−0·25
33. Shaw Test: No. of 'C' Responses	0·48	0·06	−0·20	−0·29	0·37	0·29	0·33
34. Shaw Test: No. of 'D' Responses	0·52	−0·07	−0·42	−0·46	0·30	0·15	0·47
35. Shaw Test: Total Time	0·16	0·72	−0·17	−0·42	0·64	0·53	−0·35

37. Nufferno Level Test	*−0·64*	*0·50*	*−0·09*	*−0·13*	*−0·06*	*−0·12*	*−0·80*
38. Level Test Persistence Score	*−0·13*	*0·76*	*0·18*	*−0·05*	*0·47*	*0·49*	*−0·62*
39. Nufferno Speed Test A1 Unstressed	*0·71*	*0·15*	*0·35*	*0·16*	*0·59*	*0·70*	*0·39*
40. Nufferno Speed Test B1 Unstressed	*0·48*	*0·30*	*0·38*	*0·18*	*0·54*	*0·65*	*0·12*
41. Nufferno Speed Test A2 Stressed	*0·68*	*0·18*	*0·33*	*0·13*	*0·59*	*0·68*	*0·35*
42. Word Association Test: No. of Synonyms	−0·07	−0·22	0·11	0·18	−0·21	−0·16	0·09
43. Word Assoc. Test: Repetition of Stimulus Word	0·53	0·16	0·00	−0·14	0·48	0·46	0·28
44. Word Assoc. Test: Multiple Responses	0·16	−0·02	−0·36	−0·36	0·10	−0·03	0·17
45. Word Assoc. Test: Aside Remarks	0·06	0·13	−0·43	−0·45	0·14	−0·02	−0·00
46. Word Assoc. Test: Repetition of Previous S or R	−0·08	0·23	−0·21	−0·26	0·12	0·03	−0·19
47. Babcock Motor Speed Tests	0·72	−0·06	0·11	−0·00	0·44	0·48	0·56
50. Babcock 'Substitution' Test	0·74	0·16	0·31	0·11	0·62	0·71	0·41
51. Time to Draw three squares	0·59	0·19	−0·06	−0·22	0·54	0·50	0·31
52. Wechsler Arithmetic sub-test	−0·52	0·44	−0·31	−0·34	−0·02	−0·15	−0·65
53. Wechsler Picture Arrangement sub-test	−0·52	0·25	−0·05	−0·04	−0·17	−0·20	−0·55
54. Drawing Design Test Disproportionality	0·13	−0·48	−0·00	0·05	−0·27	−0·27	0·47
55. Drawing Design Test: Average Time	0·78	0·08	−0·00	−0·16	0·58	0·57	0·52
56. U.S.E.S. Man. and Finger Dexterity	−0·51	−0·01	−0·21	−0·10	−0·35	−0·42	−0·35
57. Proverbs Test 'Abstract' Score	−0·60	0·22	0·04	0·07	−0·24	−0·23	−0·59
58. Proverbs Test: Average No. of Words	0·35	0·44	−0·41	−0·58	0·56	0·38	0·00
59. Proverbs Test: Average Reaction Time	0·31	0·30	0·36	0·19	0·43	0·54	−0·01
60. Proverbs Test: Average Total Time	0·51	0·68	0·01	−0·30	0·86	0·80	−0·08
61. Pathway I: Time Required	0·56	0·12	0·22	0·07	0·46	0·52	0·31
62. Pathway I: Errors Made	0·27	0·03	0·12	0·06	0·20	0·24	0·17
63. Pathway II: Time Required	0·57	0·13	0·28	0·12	0·48	0·56	0·31
64. Pathway II: Errors Made	0·42	−0·07	−0·10	−0·14	0·23	0·20	0·37
65. Shaw Test: Average 'A + B' Response Time	0·80	0·15	−0·00	−0·19	0·65	0·63	0·49
66. Shaw Test: Total Number of Responses	−0·26	0·53	−0·51	−0·60	0·22	0·01	−0·50
68. Static Ataxia: Maximum Backwards Sway	0·19	0·07	−0·10	−0·15	0·18	0·14	0·10

three unrotated factors of a centroid analysis would approximate closely to the first three latent vectors obtained from a principal component analysis. The factor saturations on these three unrotated centroid factors are presented in the first three columns of Table 1.103.

These three centroid factors, combined, by no means account for all the variance in the matrix of intercorrelations. Factor I accounts for 23·7%, factor II 8·99% and factor III 6·24% of the variance. Jointly, 39% of the variance is accounted for. This suggests that there are other important group factors which have been ignored. However, it was not the purpose of the present study to investigate these more restricted sources of co-variance, since they were not believed to be associated with either of the two major aspects of thought disorder being studied.

The unrotated centroid factors can, of course, be given no psychological meaning. The first factor obviously defines a sort of 'average' measure of 'overinclusion', 'retardation' and 'general intelligence' combined. Its composition is determined by the relative number of each type of test included, which was completely arbitrary. The method used to demonstrate that the three hypothesized factors could be produced by the appropriate rotation of these centroid factors will be discussed in a later section.

(*b*) *The Discriminant Function Analysis of the Factor Scores.* This section of the analysis was devised and carried out by Mr. P. Slater and is described in detail in his chapter. It is proposed to discuss mainly its psychological implications here. It was assumed that the three centroid factors extracted were linear functions of 'overinclusion', 'retardation' and 'general intelligence'. That is to say, three dimensions representing these three variables could equally well define the same three-dimensional space as is defined by the first three centroid factors. A major purpose of the present study was to discover whether or not neurotics as a group suffer from any degree of thought disorder. If they do not, none of these three factors should significantly differentiate the neurotics from the normal group. That is to say, the neurotic groups' mean points should occupy the same part of the three-dimensional space defined by the three centroid factors, as the normal group's mean point.

In order to determine this, the two neurotic groups were first compared with each other. Factor scores on the first three centroid factors were calculated for all the groups. (This was done exactly, the regression of each test on each factor being taken into account—see chapter by Slater.) A discriminant function was then computed for the two neurotic groups. This function was by definition, that combination of the three factor scores which would maximize the differentiation between the two neurotic groups. Using the equivalent of an analysis of variance technique (comparing the between groups variance to the within groups variance for this discriminant function), the two neurotic groups were found not to be significantly different. We can conclude from this, that the two neurotic groups are not significantly different on any of the three centroid factors.

The two neurotic groups were then combined, and a discriminant function (of the three factors scores) computed, which would maximize the differentiation between the combined neurotic group, and the normal group. Again, the difference was insignificant. We must conclude that the neurotic group does not differ significantly from the normal group on any of the three centroid factors. This suggests that in terms of the test variance accounted for by these three factors, there is no evidence of thought disorder in the two neurotic groups.

The normal and the neurotic groups were then combined to form a 'non-psychotic' group. The two psychotic groups were also combined. A discriminant function (of the three factor scores) was computed to maximize the discrimination between this 'non-psychotic' and 'psychotic' group. This time, the difference was highly significant, the resultant variance ratio being 13·86, significant at well beyond the 1% level. In other words, these three factors taken in conjunction, differentiate the psychotics from the remaining subjects at a high level of significance. This demonstrates that one or more of the first centroid factors is measuring some function which significantly differentiates the psychotics, as a group, from the normals and the neurotics.

Finally, the two psychotic groups were compared with each other. A discriminant function (of the three factor scores) was computed to maximize the discrimination between the schizophrenic and the depressive group. Again, the difference was significant, the variance ratio being 5·88, significant at beyond the 1% level. We can conclude again that one or more of the first centroid factors measures some function which significantly differentiates the schizophrenics from the depressives.

This analysis so far, has not demonstrated that the function which differentiates the psychotics from the other groups is not the same function which differentiates between the two psychotic groups. In other words, it is conceivable that the same dimension is involved, the 'non psychotics', the schizophrenics and the depressives all lying along a line. Such a picture would suggest that the two psychotic groups have the same sort of thought disorder, but differ only in the amount. It would be consistent with the theory that a single factor of 'psychoticism' is sufficient to account for the differences between the groups on this battery of tests. The next step of the analysis was designed to attempt to disprove this hypothesis.

(c) *The Canonical Analysis of the Discriminants.* This section of the analysis was also devised and carried out by Mr. P. Slater and is described in detail in his chapter. Only a summary will be given here. The purpose of the analysis was to discover how many uncorrelated dimensions are needed to account for the differences between the groups on the three factors. Do the five groups all lie along the same dimension (a single factor which might be called 'psychoticism'), or is more than one orthogonal dimension needed to account for the differences between the groups? We have already seen that the neurotics and the normals are not significantly different, so that we cannot expect to find that more than two orthogonal dimensions are needed to account for the differences between the 'non-psychotics', the schizophrenics and the depressives.

The procedure employed analyses the 'between groups' dispersion on the three factor scores. A principal component analysis is applied to this 'between groups' dispersion to determine how many significant latent roots (with associated latent vectors) can be found in it. In other words, does all the variation between groups lie along the same line (are all the group means in a line in this three-dimensional space?), or are two orthogonal dimensions needed to describe the 'between groups dispersion'? (must two unrelated dimensions be used to describe the positions of the group means in this three-dimensional 'factor score' space?)

The principal component analysis yielded three latent roots. (More than three would have been impossible, since there are only three factor scores.) Their values were 0·6995, 0·1763 and 0·0118. This suggests that, as far as the groups are concerned, most of the variation between the group means lies on one dimension. Its associated latent root (0·6995) is significant at beyond the 1% level. However, there is a significant amount of variation on a second, orthogonal dimension. Its associated latent root (0·1763) is also significant at the 1% level. As expected (since the neurotics and normals do not differ significantly), the third dimension contains no significant variation. Its associated latent root (0·0118) is not significant.

We must conclude, then, that a single factor of 'psychoticism' is insufficient to account for the differences between the groups on the three factor scores. Two orthogonal 'factors' are required. It will be demonstrated later that these two orthogonal dimensions are a function of one factor which can be labelled 'overinclusion' and a second factor which can be labelled 'retardation'. This analysis, of course, does not disprove the existence of the hypothesized third factor of 'general intelligence'. Since the groups were matched in this respect, such a factor could not account for differences between the groups, and could not have emerged in this analysis.

(*d*) *The Rotation and Identification of the Factors.* It was predicted from the theory of 'overinclusion' that it should be possible to extract from the matrix of intercorrelations a factor which could be labelled 'overinclusion', and that this factor should differentiate the schizophrenics from the other groups. There are two ways of testing such a prediction. The centroid factors could be rotated in such a way that the theoretically 'pure' measures of 'overinclusion' have maximum loadings on one of the factors. This factor could then be labelled one of 'overinclusion', and it should significantly differentiate the schizophrenics from the other groups. In fact, this method was attempted first. A series of rotations was performed, in the hope that such a factor could be achieved. This proved almost impossible. With so many variables to take into account, it was extremely difficult to imagine what would happen to them all, before each rotation was performed. The imaginary manipulation of fifty-four variables in a three-dimensional space is not easy. Although rotating a factor through a single test, or even a small and obvious cluster is simple enough, no method of rotation whereby a specified number of tests can be made to have a maximum joint loading on a single factor was then known to us. This approach to the problem was, therefore, abandoned.

The alternative method is to obtain that factor which maximizes the discrimination between the schizophrenics and the other groups. According to the theory, this should be a factor of 'overinclusion', and the theoretically 'pure' measures of overinclusion should have the highest loadings on the factor if the hypothesis is correct. It was this method which was used.

The findings described earlier suggested that both the schizophrenics and the depressives were retarded. The schizophrenics, however, were also overinclusive. It was thought that this would prove to be an unrelated abnormality. Therefore, it was predicted that that factor which maximized the discrimination between the depressive and schizophrenic group, would be a factor of 'overinclusion', and that the theoretically 'pure' measures of overinclusion would have the highest loadings on it.

The technical procedure used was suggested by P. Slater. The factor was obtained quite simply, by utilizing the data obtained in the discriminant function analysis. A discriminant function of the three centroid factor scores had been obtained which maximized the differentiation between the schizophrenic and the depressive group. Thus, the test loadings on the desired factor were obtained merely by pre-multiplying the discriminant function weights (for centroid factors I, II and III) into the row of factor saturations (for centroid factors I, II and III) for each test.[1] The resultant figure gives the saturation of the test on the factor which differentiates maximally between the schizophrenics and the depressives. This set of saturations is shown in Table 1.103 under the column headed 'Factor O'. This factor was regarded as an 'overinclusion' factor. The mean factor scores for this factor are shown in Table 1.104. As can be seen, only the Schizophrenics have extreme scores, the other group means being nearly identical.

The findings already described, suggested that both psychotic groups were retarded. It was thus predicted that a factor of 'Psychoticism'

---

[1] These weights are as follows: factor $I = -0{\cdot}1718$, factor $II = -0{\cdot}3196$, factor $III = 0{\cdot}9318$.

TABLE 1.104

Mean Factor Scores for the Groups on the 'Overinclusion' Factor

Normals:	− 10·246
Dysthymics:	− 11·463
Hysterics:	− 12·761
Depressives:	− 11·215
Schizophrenics:	− 16·568

Note: These mean scores were obtained by pre-multiplying the mean scores for each group on the three centroid factors (see chapter by Slater for these scores) by the weights obtained in the discriminant function analysis for differentiating the schizophrenics from the depressive group.

would in fact be a factor of 'slowness' or 'retardation'. This factor was obtained in a similar way. It was defined as that factor which differentiated maximally between the mean of the two psychotic groups combined, and the mean of the normal and neurotic groups combined. Again from the discriminant function analysis described, a set of weights had

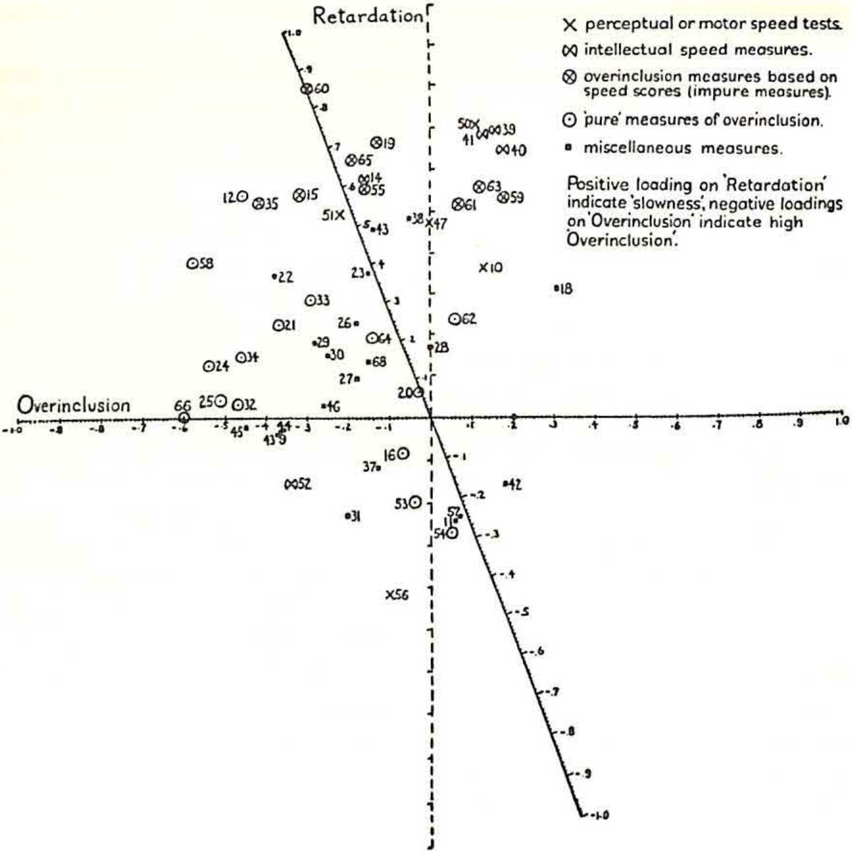


FIG. 1.7. The Factors of 'Retardation' and 'Overinclusion'.

already been obtained to apply to the three unrotated factors to produce this discriminant function.[1] To obtain the test saturations on this factor, these weights were pre-multiplied into each row of test saturations (for the three centroid factors). The resulting figure gives the saturation for each test on this 'psychoticism' factor. The test saturations on this factor are given in Table 1.103 under the column headed 'factor P'. These two factors are reproduced geometrically in Fig. 1.7. The 'overinclusion' factor is marked on the diagram. The 'psychoticism' factor is marked 'retardation' and is represented by the heavy black line. These two factors ('Overinclusion' and 'Psychoticism') are not uncorrelated. The angle between them is 69° 10′. The correlation between these two factors is $-0{\cdot}484$ using the formula:

$$r_{iK} \text{ (the correlation between two vectors 'i' and 'K')} = \frac{i'MK}{\sqrt{i'Mi\,.\,K'MK}} \quad \text{where M = the original matrix of correlation coefficients.}$$

Although this 'psychoticism' factor differentiates maximally between the psychotics as a group, and the combined normal and neurotic group, because it is correlated with the 'Overinclusion' factor, it also differentiates between the schizophrenic and the depressive group, the schizophrenics being more 'psychotic' (see Table 1.105). This factor was, therefore, rotated so that it had two properties: (1) it minimized the differentiation between the two psychotic groups,[2] and (2) it now

TABLE 1.105

Mean Factor Scores for the Groups on the 'Psychoticism' Factor

Normals:	16·905
Dysthymics:	21·774
Hysterics:	20·900
Depressives:	28·268
Schizophrenics:	33·045

Note: These mean scores were obtained by pre-multiplying the mean scores for each group on the three centroid factors (see chapter by Slater for these scores) by the weights obtained in the discriminant function analysis for differentiating the psychotic from the non-psychotic group.

---

[1] These weights are: factor I = 0·6736, factor II = 0·7390, factor III = − 0·0056.

[2] The two psychotic groups will still not be exactly equal in terms of their factor scores. In a discriminant function analysis, a vector which maximizes the difference between the means of the groups is obtained *when each within group dispersion has been made spherical (i.e. each group has unit variance in every dimension)*. This procedure should not, strictly speaking, be used when the groups differ significantly in their dispersion. When the groups differ with respect to their dispersion as in the case of the present study (where the schizophrenics almost certainly have a larger standard deviation on all three factors), the discriminant function only differentiates maximally between these 'standardized' scores with unit variance in all dimensions. However, these slight inaccuracies are probably not important.

passed through the mid point of the cluster of speed tests. This rotation was effected by putting it at 90 degrees to the factor of 'overinclusion'. This rotated factor can be called a factor of 'psychoticism' since it still differentiates the psychotics from the other groups. Indeed it is, perhaps, a better factor of 'psychoticism' since both schizophrenics and depressives now have almost equally abnormal scores on it. However, an inspection of Fig. 1.7 makes it apparent that this factor is now one of 'retardation', as it is defined by a large group of speed tests. This 'retardation' factor is shown in Fig. 1.7 by a broken line at right angles to the 'overinclusion' factor. The loadings of the tests on the factor of 'retardation' are shown in Table 1.103 under the heading 'Factor R'. The correlation between the 'Overinclusion' factor and the 'Retardation' factor is now only −0·377 so that they are more nearly independent.[1]

The third factor, which has been labelled 'General Intelligence' was defined as that factor which is at right angles to the plane defined by the 'Overinclusion' and the 'Psychoticism' (or the 'Overinclusion' and 'Retardation') factors. In order to obtain the test saturations on this factor a set of three weights was computed which defines a vector at right angles in the three-dimensional space to both the set of discriminant function weights from which the 'O' factor derived, and the set of discriminant function weights from which the 'P' factor derived.[2] The resultant set of three weights was then pre-multiplied into the row of factor saturations for the three centroid factors, for each test. The resultant figure gave the saturation on this 'G' factor for each test. These saturations are presented in Table 1.103 under the heading 'Factor G'. The 'G' factor is thus by definition that factor which does *not* differentiate the psychotic groups from one another, or from the combined neurotic and normal group. Since we know that the neurotics do not differ significantly from the normals, the 'G' factor cannot differentiate any of the groups significantly. This is as it should be, since the groups were matched for pre-illness intelligence level.

(*e*) *The Make-Up of the Factors*

The 'Overinclusion' Factor. It has already been shown that the 'overinclusion' factor differentiates significantly (and maximally) between the schizophrenics and the depressives. An inspection of Fig. 1.7, makes it clear that as expected, most of the 'pure' measures of overinclusion have high loadings on this factor. Nine out of the fifteen relatively 'pure' measures of overinclusion have loadings on this factor greater than 0·25.

---

[1] Even though these factors are at right angles, they are still correlated. The only completely uncorrelated set of factors set at right angles, are the original, unrotated centroid factors—or more precisely, the original principal component vectors. However, as was pointed out, these have no psychological meaning.

[2] These weights are: for factor I = + 0·7354, for factor II = − 0·6710, for factor III = − 0·0945.

The eight highest loadings on this factor are obtained by tests of 'overinclusion'. Furthermore, all eight were regarded as 'pure' measures except one (Variable 35—total time for the Shaw test) which, it was expected, would also be a measure of slowness (as is the case). These results appear to confirm the theory that these tests do indeed measure 'overinclusive' thought disorder in common, and that this disorder is characteristic of schizophrenic patients.

Of all the twenty-three measures of 'overinclusion' included in the factor analysis, whether or not they were regarded as 'pure' measures, only four produced loadings on the 'overinclusion' factor opposite to prediction. These four tests had very small loadings indeed. The average 'reaction time' on the proverbs test (Variable 59) had a loading of +0·19, opposite to prediction. (Note that a negative loading indicates the 'overinclusive' end of the 'overinclusion' factor.) Thus, overinclusive subjects tend to have a slightly faster reaction time, in spite of the fact that they talk at greater length and use more words. The total time taken for Pathway I had a loading of +0·07, and the number of errors on Pathway I had a loading of +0·06, both opposite to prediction, but both very small indeed. It will be recalled that neither score differentiated the groups significantly. It would appear, as suggested earlier, that this test is too simple to be a measure of 'overinclusion'. The other score which had an unpredicted loading, was the time taken for Pathway II. It had a loading of only +0·12. Again, this score did not differentiate the groups significantly.

It was expected that the 'concreteness' ratings for the Goldstein-Scheerer Colour-Form test (Variable 22), the Goldstein-Scheerer Object-Sorting Test (Variable 23), and the Benjamin Proverbs test (Variable 57) would be partly a function of 'overinclusion', because schizophrenics are rated 'concrete' on the basis of their unusual responses. This prediction was verified for each test. The Colour-Form test has a loading of −0·38 ('concrete' scores being more 'overinclusive'), as expected. The Object-Sorting test rating has a loading of −0·15 ('concrete' scores being more 'overinclusive'), and the Benjamin Proverbs test rating has a loading of +0·07 ('abstract' scores being less 'overinclusive') as expected.

Several of the miscellaneous tests produced interesting saturations on the factor of 'overinclusion'. Overinclusive subjects on the word association test produce fewer synonyms (loading = +0·18), more often repeat the stimulus word (loading = −0·14), give more 'multiple responses' (loading = −0·36), give more 'aside remarks' (loading = −0·45), and more often repeat a previously given stimulus or response word (loading = −0·26). This was exactly as predicted, on the grounds that overinclusive subjects would generally tend to produce more 'distant' or 'unusual' responses, although none of these scores differentiated the groups at an acceptable level of significance.

It is also of interest, that on the Luchins test, before they have been given the 'set', overinclusive subjects more frequently use the 'simple' method of solution (loading = +0·31). While this measure did not differentiate the groups significantly, it throws some doubt on the finding that schizophrenics tend not to develop the 'set' as expected from the theory. It is possible that they went into the test situation with an undue preference for the simple method even before the 'set' had been given, although why this should be it is difficult to understand. One would expect that, if anything, 'overinclusive' people should prefer the more complex (three-jar) solution to the simpler (two-jar) solution before the 'set' is given.

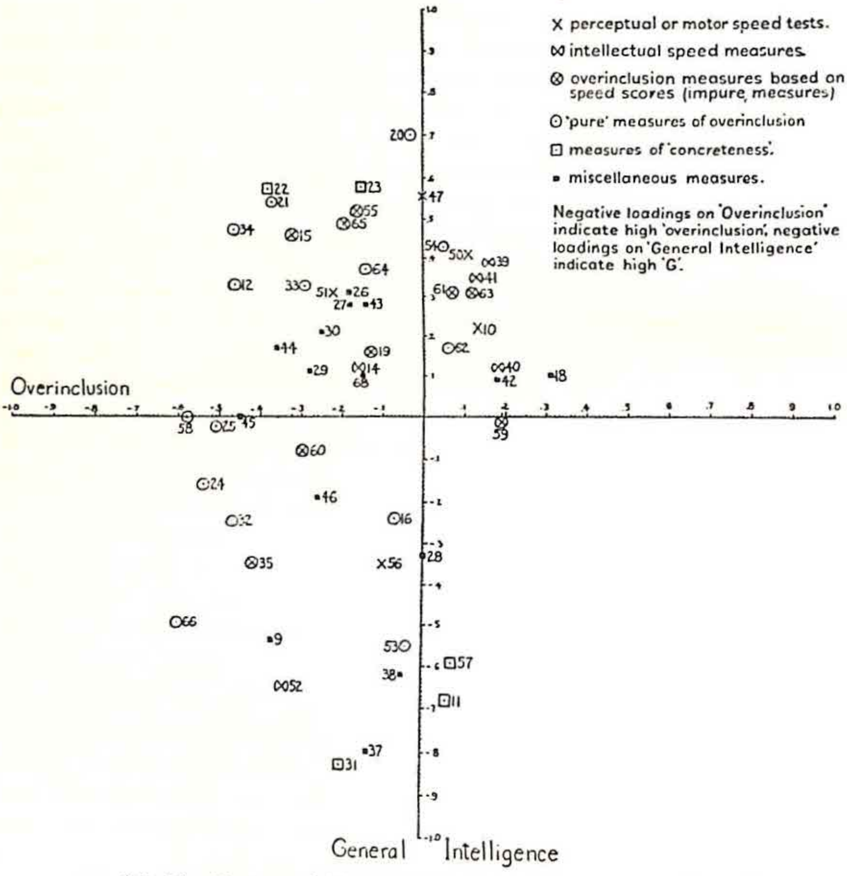


FIG. 1.8. The Factors of 'Overinclusion' and 'General Intelligence'.

For some reason or other, even though the groups were matched for vocabulary, overinclusive subjects tend to have higher Mill Hill Vocabulary scores (loading −0·37).

As might be expected, most of the measures of 'overinclusion' are also a function of General Intelligence. Fig. 1.8 shows the results obtained when the 'overinclusion' and 'G' factors are plotted together.

On some tests, intelligence is positively correlated with overinclusion. For example, both intelligent and 'overinclusive' people give a larger number of 'B' responses to the Shaw test, as well as a larger total number of responses. On other tests, the correlation is negative. For example, both 'overinclusive' and duller people give a large number of 'non-A' responses on the Object Classification test, and a large number of 'C' and 'D' responses on the Shaw test. Only two of the best measures of 'overinclusion' are completely independent of intelligence. These are the average number of words per proverb (Variable 58) and the number of unusual responses given on the Goldstein Object-Sorting test (Variable 25). However, it is obviously possible to assemble a battery of 'overinclusion' tests so that the effect of intelligence is balanced out.

The factor of overinclusion was defined as that factor which differentiated maximally between the depressive and schizophrenic groups. This does not preclude the possibility that depressives are, nevertheless, more overinclusive than normals or neurotics. In Tables 1.104 and 1.105, the mean factor scores for all the groups have been shown for the 'Overinclusion' and the original 'Psychoticism' factor. It can be seen that, as a group, the depressives are no more 'overinclusive' than the normals and the neurotics.

While it is possible to compute precisely the mean factor scores for the different groups, it is extremely laborious to calculate the other parameters of the distributions of each group (see chapter by Slater). In order to estimate the range of variation on these factors, and in order to see how well a simple, practical combination of the tests would measure the factors and differentiate the groups, simple 'factor scores' were derived for both 'Overinclusion' and 'Retardation'. No attempt was made to weight any of the tests differentially. However, care was taken to 'balance' out the influence of the other factors in the construction of each factor score. The factor score for 'overinclusion' consisted merely of the sum of the transformed scores for the following tests: (This does not even weight the tests exactly equally. By and large the transformed scores have similar standard deviations, but this is not exactly the case. Standardized scores could have been used, although this is a more laborious procedure.)

Variable 12: Object Classification test 'non-A' responses.
Variable 16: Luchins test: Number of 'simple' solutions following 'set'.
Variable 21: Epstein test: Neologism score.
Variabe 24: Goldstein Object-Sorting test 'unusual' responses.
Variable 25: Goldstein Object-Sorting test 'Overinclusion' score.

Variable 31: Shaw test number of 'A' responses.
Variable 32: Shaw test number of 'B' responses.
Variable 34: Shaw test number of 'D' responses.
Variable 37: Nufferno Level Test Score.
Variable 53: Wechsler Picture Arrangement Test score.
Variable 56: USES Combined Manual and Finger Dexterity score.
Variable 58: Average Number of Words per proverb.
Variable 66: Shaw test total responses.
Minus Variable 42: Word Association test number of synonyms.
Minus Variable 57: Proverbs Test 'Abstraction' score.

The results obtained with this approximate factor score are shown in the histograms in Fig. 1.9. It can be seen that the depressives on this score are all completely within the normal range. The means for the groups are as follows: Normals, 25·75, Dysthymics 24·20, Hysterics,

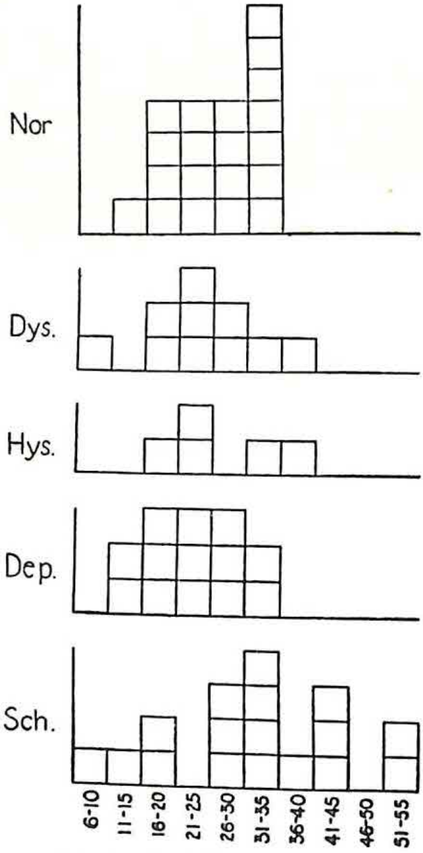


FIG. 1.9. Overinclusion Factor Scores.

27·40, Depressives, 23·08, Schizophrenics, 32·12. (Note that not all the subjects are included. Subjects with incomplete scores were omitted, instead of making the appropriate mean adjustment as in the case of some other calculations.) The schizophrenics on this score are significantly differentiated from all the other groups combined, at beyond the 1% level. While there is considerable overlap for this score, it must be remembered that no attempt was made to weight the tests in accordance with the extent to which they measure the factor. A more accurate factor score would have differentiated the groups better.

The Factor of 'Retardation'. This factor has two properties. It differentiates between the psychotics as a group, and all the other subjects, but it does not differentiate between the schizophrenics and the depressives. It was expected that this factor would be a general measure of 'retardation', and indeed it was rotated into this position (at right angles to the factor of 'overinclusion') partly because it would then pass through a large cluster of speed tests. The highest ten loadings on this factor are all speed scores of one sort or another. Twenty-five scores have loadings greater than ±0·30 on this factor. Only six of these are not speed scores.

The most striking characteristic of the 'retardation' factor, is that it is perfectly general. There is no tendency for any special type of speed measure to have the highest loading. It is thus an average measure of intellectual, perceptual and motor speed, and can quite fairly be labelled a factor of 'retardation'. The measures of overinclusion, based on speed scores, tend, of course, to have the predicted loadings on the factor of 'overinclusion', but these speed scores have as high loadings on the factor of 'retardation' as any of the other speed scores. Fig. 1.7 shows the factor of 'retardation' plotted against the factor of 'overinclusion'. Positive loadings indicate a 'slow' performance. All the speed test loadings on the 'retardation' factor are consistent. The few positive loadings in Fig. 1.7 result from speed scores in which a fast performance is given a high score. (On most of the measures, a slow performance is given a high score.)

While a detailed theory was elaborated to explain the fact that measures of 'overinclusion' define a common factor, no such theory has been put forward to account for the factor of 'retardation'. It was initially supposed that schizophrenics would not prove to be generally retarded. Depressives were expected to be generally retarded on the grounds that their depressive preoccupations would interfere with their performance on all types of test, producing a general distraction. It was supposed that schizophrenics would be slow only on intellectual tests, as a 'by product' of their overinclusion. This prediction is clearly not the case. Schizophrenics are generally retarded on all types of speed test, and, as can be seen from Fig. 1.7, intellectual speed tests are no more highly saturated on the 'overinclusion' factor than are motor and perceptual speed tests. Nevertheless, the 'retardation' factor and the 'overinclusion' factor correlate to the extent of −0·377. It was not possible to produce a factor of retardation with a lower correlation with the factor of 'overinclusion' defined. This is hardly surprising. In the original matrix of correlations, the speed tests tended to correlate with the measures of overinclusion, and it is not therefore possible to produce any linear combination of overinclusion scores which will be completely independent of a linear combination of speed scores. Why overinclusion tends to produce a slight general retardation is not entirely clear. It is possible, however, that 'overinclusive' individuals tend to over complicate the instructions of even the simplest motor tests, paying attention to irrelevant details, and thus retarding their performance. In short, 'overinclusion' is completely unrelated to the level of complexity of the task.

This fairly low correlation between 'overinclusion' and 'retardation' nevertheless cannot entirely explain the fact that the schizophrenics, on average, show as extreme a degree of general retardation as do the depressives. Some other possible explanations for this will be considered in the final section.

An approximate factor score was calculated for the factor of 'retardation', in order to estimate the ranges of variation on this factor for the different groups. This score consisted merely of the sum of the transformed scores for the following tests:

Variable 10: Brengelmann Picture Recognition test.
Variable 14: Object Classification test average response time.
Variable 19: Epstein test total time.
Variable 39: Nufferno Speed test A1.
Variable 40: Nufferno Speed test B1.
Variable 41: Nufferno Speed test A2.
Variable 47: Babcock Motor Speed tests.
Variable 50: Babcock 'Substitution' test average time per line.
Variable 55: Drawing Design test total time.
Variable 59: Reaction Time, Proverbs test.
Variable 60: Proverbs test, average total time.
Variable 61: Time required for Pathway I.
Variable 63: Time required for Pathway II.
Variable 65: Shaw test average 'A + B' response time.
Minus Variable 56: USES Manual and Finger Dexterity score.

The results for this factor score are shown in Fig. 1.10. Note that, as with the comparable 'O' factor scores, some of the cases have been excluded for the same reason.

The means for this score are as follows: Normals, 28·40; Dysthymics, 37·10; Hysterics, 31·20; Depressives, 50·38; Schizophrenics, 52·12. The two psychotic groups combined differ significantly from the remaining subjects at beyond the 1% level. This factor score for the 'R' factor correlated −0·139 with the similar factor score for the 'O' factor ('overinclusive' people being slightly faster). This result is in the opposite direction to that obtained when the exact correlation between 'O' and 'R' was calculated. This is almost certainly because these factor score estimates are not based on all the tests, and because the tests used are not weighted in accordance with their factor saturations. Too much attention should, therefore, not be paid to this less accurate correlation coefficient. However, it does emphasize the fact that the 'O' factor and the 'R' factor are relatively independent, especially when only the best measures of each are taken into account.

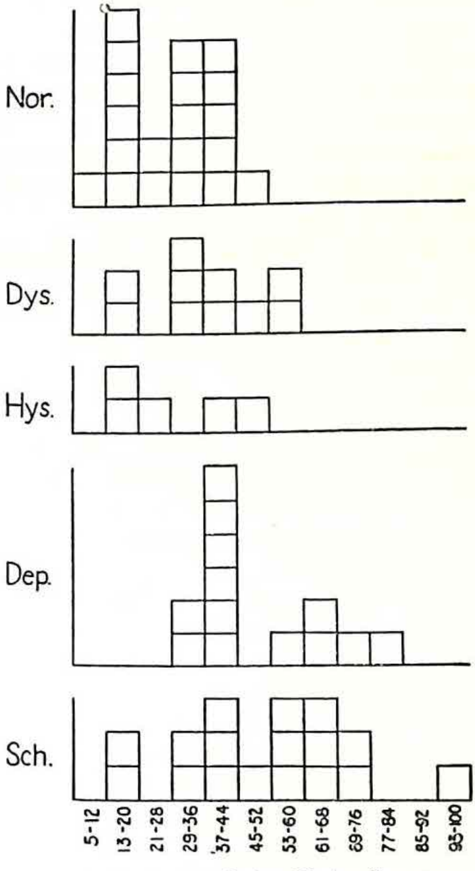


FIG. 1.10. Retardation Factor Scores.

It is interesting that although they have the same mean score, the depressives are more homogeneous on this score than are the schizophrenics. More schizophrenics are within the normal range than are depressives. Of the three non-psychotic groups, the dysthymics have the highest mean scores. The dysthymic group is of some interest, since eight of the twelve members of the group had depressive symptoms, seven being diagnosed as 'reactive depression'. It is not surprising that they tend to be similar to the 'endogenous depressives'. Indeed, it is surprising that speed scores differentiate these two groups as well as they do, especially in view of the fact that many of the doctors who selected these cases believed that the 'endogenous'-'reactive' distinction was a difficult one to make. These findings throw some doubt on the notion that depressed patients are generally retarded because they are distracted by their depressive thoughts. Both 'endogenous' and 'reactive' depressives were equally depressed in this respect. To some extent, the discrimination between these groups on speed tests may be contaminated. One of the diagnostic criteria used by the doctors in judging between 'endogenous' and 'reactive' depression in the present study was the presence of 'retardation' which was judged an 'endogenous' feature. It is not surprising, perhaps, that speed tests should, therefore, differentiate the groups so well. However, the fact remains that a large group of depressed patients is extremely retarded on all measures of speed of functioning, and the present study throws little light on the cause of this. It might well have nothing to do with the worries and preoccupations of the patients, since these features are common to all the depressives. It may have some underlying physiological correlates, and these may be the main difference between the 'endogenous' and 'reactive' groups. However, one can only speculate about what these physiological differences might be.

The Factor of 'General Intelligence'. This factor is very clearly defined indeed. On this factor, a negative loading indicates a positive correlation with intelligence, a positive loading, a negative correlation (see Fig. 1.8). Among the highest negative loadings are the Nufferno Level Test (−0·80), the Mill Hill Vocabulary IQ (−0·54), the Wechsler Arithmetic sub-test (−0·65) and the Wechsler Picture Arrangement sub-test (−0·55). Furneaux has found that intellectual speed, as measured by the Nufferno speed tests, has a correlation of about 0·40 with 'level' (which might be identified with 'G' in the present context). Nufferno Speed test A1 has a loading of +0·29, Nufferno test A2, a loading of +0·35, and Nufferno test B1, a loading of +0·12, almost exactly what we would expect to find on a factor of General Intellectual Level. (These loadings are positive, as log time scores were used in the correlations. A large score indicates a poor intellectual speed.)

The most striking results obtained on this factor, are the loadings obtained by the five measures of 'concreteness'. It was expected that these tests would to some extent measure 'overinclusive' thinking (because of the ratings of 'concrete' conventionally applied to some unusual methods of sortings produced by the schizophrenics), but would otherwise be measures of general intellectual level. Since this variable was controlled, their 'G' variance would not, of course, differentiate the groups. This was very clearly borne out by the factor analysis. The five measures of concreteness all have very high loadings on the 'G' factor as follows: Goldstein-Scheerer Colour Form 'concreteness' score, +0·57; Goldstein-Scheerer Object-Sorting test 'concreteness' score, +0·58; Benjamin Proverbs test 'abstractness' score, −0·59; Shaw test number of 'A' responses (normal, 'abstract' responses), −0·83; Object Classification test number of 'A' responses (normal, 'abstract' responses), −0·68. These loadings suggest that these scores are just as good intelligence test items, as are the two Wechsler Bellevue sub-tests included.

One other finding is of interest in connection with this factor. The most disappointing measure of 'overinclusion' used in the present study, was the Epstein test 'overinclusion' score. This score did not differentiate the groups significantly, and had a very low loading on the factor of 'overinclusion'. However, it has a loading of +0·70 on the 'G' factor. This suggests that on the whole, it is dull people who underline more words on this test (and to a very small extent, 'overinclusive' people). This is almost certainly a function of word knowledge. Dull people, who do not know what all the words mean, probably tend to guess, and thus obtain an unusually high 'overinclusion' score. If this is the case, the Epstein test could only be used as a measure of overinclusion with very bright people who are certain to have learned the meanings of all the words prior to the onset of their illness. While the 'neologism' score is a better measure of 'overinclusion', dull people also tend to underline neologisms, presumably not realizing they are not real words (loading = +0·54).

VIII. SUMMARY AND DISCUSSION

The present research investigated three main aspects of thought disorder in psychotic patients, abnormal slowness of psychological function or 'retardation', 'overinclusion', and 'concreteness'. A group of tests of 'psychoticism' was also included, but of those which differentiated the groups significantly, all but one was a test of perceptual or motor speed.

Five groups were tested, twenty normal people, twelve 'dysthymic' neurotics, eight hysterics, twenty endogenous depressives, and twenty schizophrenics. The groups were carefully matched on measures of pre-illness general intelligence, age and education.

It was found that most of the tests which were thought to measure 'overinclusive' thinking, intercorrelated significantly. Nearly half the tests discriminated the groups significantly. In each case, it was the schizophrenic group which obtained abnormally high scores, the other groups getting low scores. A factor of 'overinclusion' was obtained from an analysis of the matrix of intercorrelations. This factor was defined as that rotated centroid factor which maximized the differentiation between the depressive and the schizophrenic groups. It was shown that most of the tests of 'overinclusion' had high loadings on this factor. Factor scores were derived for each group. Again, only the schizophrenic group obtained an extremely 'overinclusive' mean score. The depressives were no more overinclusive than the neurotics or the normal group. It could thus be concluded that 'overinclusive thinking' is an abnormality which characterizes only schizophrenic patients.

The tests of slowness of psychological function also correlated highly, and produced a well-defined factor which was labelled 'retardation'. This factor had a low correlation with the factor of 'overinclusion', 'overinclusive' subjects tending to be slightly retarded. It was obtained in such a way that it differentiated the two psychotic groups significantly from all the other subjects. However, it did not differentiate significantly between the schizophrenics and the depressives. The striking feature of this factor is that it is measured by all the speed tests, regardless of their content. It is a combined measure of intellectual, motor and perceptual slowness, and was thus labelled 'retardation'. It had initially been thought that slowness in schizophrenia was confined to intellectual tasks, as it was a by-product of 'overinclusive' thinking in this group. This hypothesis was clearly disproven. 'Overinclusive thinking' produces a slight amount of retardation on all tests, regardless of content. Quite unexpectedly, however, the schizophrenic group is just as retarded on average as the depressive group. This extreme degree of retardation cannot be explained as due to 'overinclusion'.

Three of the individual measures of 'concreteness' differentiated the schizophrenic group from the other groups. It was shown that this was due to the fact that on these tests, when the conventional system of rating 'concreteness' is adopted, many of the unusual responses given by overinclusive schizophrenics tend to be labelled 'concrete'. Other measures of the ability to 'abstract' or 'generalize' did not differentiate any of the groups. It was shown that 'concreteness' scores are almost entirely a function of general intelligence. They would not, therefore, differentiate between any abnormal groups which are matched for general intelligence. In other words, there is no evidence at all that any psychotic illness can produce a specific disability which could be labelled 'concreteness'.

While these findings are consistent, and fairly clear as far as they go, they nevertheless raise a number of interesting problems. In the present study, the entire emphasis has been on the average characteristics of each group. In that sense, the psychiatric diagnostic label has been the main reference point. The underlying assumption has been that the labels used refer to more or less homogeneous groups. However, this is clearly not the case. Perhaps the most striking fact in the present study, is the heterogeneity of the schizophrenic group. If the histograms for the various specific measures of 'overinclusion' are examined (Figs. 1.1, p. 30, 1.2, p. 34, 1.3, p. 37, 1.4, p. 46, 1.5, p. 53, 1.6, p. 58) and for the 'overinclusion' factor score (Fig. 1.9, p. 96), it is clear that the 'schizophrenics' are so heterogeneous that it is difficult to regard them as a single group. One fact that stands out clearly, is that their range extends from the bottom of the normal range, to the highest scores obtained on most tests. On average, roughly half the schizophrenic group is within the normal range, while half the group is outside the normal range. On some of the measures, the distributions are nearly bi-modal. Thus it is clear that only about half the present schizophrenic patients are abnormally overinclusive. The other half are normal in this respect. It is highly likely that this type of thought disorder is related to a specific set of symptoms. As has been argued elsewhere (Payne, 1960) it is tempting to relate abnormal 'overinclusion' to the development of ideas of reference, and paranoid delusions. Both these symptoms could be regarded as unwarranted generalizations from the facts. Such patients often tend to perceive relationships where none exist, and to take as evidence for their views, data which most people would judge to be irrelevant. No attempt was made to investigate the relationship between test scores, and type of symptom in the present study. The cases were too few, and this is the aim of a further investigation. It was certainly the case in the present study, however, that the two individuals who obtained the most extreme scores on most of the overinclusive tests, were diagnosed as 'Paranoid Schizophrenic'. If this were the case, a further interesting problem is raised. There is evidence in the literature (Payne, 1960) that paranoid schizophrenics are less slow, and show less general cognitive deterioration than other schizophrenic sub-groups. It is not commonly found that they are as slow as a typical group of 'depressives'. Nevertheless, in the present study, the schizophrenic group was on average, just as retarded as the depressive group. However, if the histograms for the 'retardation' factor scores are examined (Fig. 1.10, p. 99) it is clear that the schizophrenics are heterogeneous on this measure, too. The 'endogenous' depressives generally appear retarded. None of them get such fast scores as the fastest normals. However, a number of the schizophrenics are as quick or quicker than the normal average, just as a number are much slower than the endogenous depressive average. It is tempting to speculate that it is the extremely overinclusive schizophrenics, who have ideas of reference and paranoid delusions, who are nearly normal with respect to speed of functioning. (Apart from slight slowness resulting from their overinclusion.) This would be consistent with the fact that these two factors are virtually uncorrelated.

In other words, it seems likely, that the 'schizophrenic' group can be subdivided into two quite distinct groups. Those with ideas of reference, or paranoid ideas, might be extremely overinclusive, but not much retarded. The remainder of the group might contain a number of schizophrenics who are not overinclusive at all, but who are much more retarded than the depressive group. It is conceivable that this sub-group might have catatonic symptoms.

These predictions were tested in part on the present data, using the 'approximate' factor scores described, and plotted in figures 1.9 and 1.10. No normal subjects obtained a 'retardation' score in excess of 48, while ten schizophrenics had higher retardation scores than this. These ten abnormally retarded schizophrenics were compared with the seven schizophrenics whose 'retardation' scores were within the normal range. The ten abnormally retarded schizophrenics had an average 'overinclusion' score of only 27·2, nearly identical with the normal mean of 25·8. Thus, the abnormally retarded schizophrenics are clearly not abnormally overinclusive as a group. On the other hand, the seven schizophrenics whose retardation scores were all within the normal range, had a mean 'overinclusion' score of 39·14. It is clearly the non-retarded schizophrenics who tend to be abnormally overinclusive in the present study.

It is obviously necessary, not merely to relate these psychological abnormalities to psychiatric labels, but to discover how they might be related to specific symptoms, and what their prognostic significance might be. In this way, objective psychological research can aid psychiatry in altering its system of classification in a way which might make the aetiology of some conditions easier to understand, and the prognosis and response to treatment more clearly predictable.

PART TWO

Experiments in Psychodynamics

THE EXCITATION-INHIBITION BALANCE IN NEUROTICS AND IN NORMALS

2

THE EXCITATION-INHIBITION BALANCE IN NEUROTICS

Gordon Claridge

INTRODUCTION

THE statement that almost any objective measure of behaviour we care to take will inevitably reveal wide individual differences in performance is so commonplace that its reiteration here may seem scarcely necessary. Yet, in their search for general laws of behaviour, psychologists have tended to minimize the importance of such variations, feeling, no doubt, that overemphasis on them may unduly restrict the generality of the principles they have enunciated. To the personality theorist, of course, it is these very individual differences which are of central importance. since any attempt to describe, classify, or measure personality is, almost by definition, itself an exercise in the analysis of individual variations in response. Only fairly recently in the history of psychology has there been, however, a realization that a useful link can be made between personality factors and previously unexplained variations in performance on tests hitherto largely outside the immediate interest of the personality theorist. Furthermore, there is now increasing confidence that the systematic study of this link between these two approaches to behaviour analysis can become a means of furthering, rather than handicapping, the search for general psychological laws.

Whichever way we look at the problem, however, whether we are primarily interested in the determinants of performance of a particular kind or in seeking objective measures of the variables underlying personality, it is clear that the isolated demonstration of a connection between a particular performance measure and personality is of limited usefulness unless it can be related to a broader theoretical framework, which should, according to its comprehensiveness, be capable of taking account of a wide variety of apparently different measures of behaviour. Probably the most comprehensive framework of this kind to date is the attempt of Eysenck (1955a) to link a number of personality dimensions, previously demonstrated by him, to some of the variables proposed by Hull (1943) to underly behaviour in general. By doing this, Eysenck has been able to account in a meaningful way for individual differences in performance using a variety of experimental techniques. In particular, his suggestion that the learning theory variable of reactive inhibition (I_R) is at least one of the factors underlying the personality dimension of introversion-extraversion has led to a number of direct tests of the hypothesis arising from this that extraverts will show a greater tendency to build up inhibition than will introverts. Similarly, according to the original dimensional analysis, neurotics of the hysteric type, being extraverted, should build up inhibition much more rapidly than neurotics of the dysthymic type, who are introverted. Tests of this general hypothesis have been made using conditioning techniques (Franks, 1956), perceptual motor tasks, particularly the pursuit rotor (Eysenck, 1956a), and short perceptual tests (Eysenck, 1955a). For a fuller account of the theory than can be given here and for experimental evidence relating to it the reader is referred, in addition to the original statement of postulates cited above (1955b), to Eysenck (1957).

Although it is certainly possible that variables other than those relating to the inhibitory process emphasized by Eysenck may ultimately be shown to be responsible for the results of some of the studies quoted by him, on the whole it is probably true to say that the original assumption does, in fact, account for most of the findings in the most parsimonious way. It was, at any rate, within this general theoretical framework and with the specific hypothesis linking extraversion to inhibition as a starting-point that the investigation reported here was planned. The principal aim of the study was, in fact, to investigate the relationship between central inhibition and extraversion in a systematic fashion, using as wide a variety of experimental measures as possible. Predictions for most of the tasks used which were relevant to the inhibition hypothesis could be made with some confidence in the case of those subjects whose personality could be adequately defined within the two dimensions of neuroticism and introversion-extraversion, viz. hysterics, dysthymics, and normals.

The inclusion of a group of psychotics introduced a secondary aim into the study, which was to investigate the performance of schizophrenics relative to that of normal and other psychiatric groups. Predictions on the basis of an inhibition hypothesis alone could be made with less certainty in the case of this group, since it was here that there was the greatest possibility that other variables would be found to be operating. The analysis of psychotic performance in previous work has, in fact, been approached from a number of different angles, the two most important of which are a formulation in terms of some kind of inhibitory process and an analysis along motivational lines. That abnormally rapid growth rates of internal inhibition may be characteristic of schizophrenics during continuous work was, indeed, a possibility considered, for instance, by Venables and Tizard (1956a), and reasserted by the same authors (1956b) following an investigation of reaction time in psychotics. This is clearly in keeping with Pavlov's original theory of 'protective inhibition' in schizophrenia (1941). However, other workers, such as Rodnick and Shakow (1940), have tended to stress the poor motivation of the schizophrenic in terms of his inability to maintain attention or 'readiness to respond' in reaction time situations. It is, of course, quite apparent that these two kinds of approach are not of necessity basically different or incapable of being handled simultaneously within the framework of contemporary learning theory, particularly if one accepts the view, which is well supported (Ammons and Ammons, 1951; Kimble, 1952), that reactive inhibition is a central, rather than, as Hull himself thought, a peripheral phenomenon. The balance between positive motivational level (D) and the central negative drive state (I_R) may be considered a critical determinant of measured performance, as the modification of Hullian theory by Kimble (1949), and Kimble and Shatel (1952), has indicated. Nevertheless, there were still, at the outset, at least two possible variables on which the schizophrenics might be differentiated—the basic level of positive drive and the degree of tendency towards negative drive, or both. It was clear, therefore, at the beginning of the investigation that a general hypothesis about the behaviour of schizophrenics was difficult, if not impossible, to formulate, particularly in view of the relatively wide variety of experimental techniques employed. For that reason it was necessary to proceed on a somewhat empirical basis as far as the psychotic group was concerned, making predictions where previous work suggested them or, if this were not possible, applying the measures in the hope that systematic findings would emerge which could be incorporated within the framework of Eysenck's behaviour theory and subsequently yield further hypotheses, which could be tested at a later date.

The choice of experimental tasks used in the investigation was governed by the desire to take a relatively wide cross-section of behaviour, covering those areas which could be roughly designated as perception, attention, and learning. Results from work in the fields of perception and learning have, as we have briefly indicated above, already lent support to the general hypothesis put forward by Eysenck of a relationship between inhibitory growth and extraversion. In fact, some of the tasks used in the present study were those employed previously in testing this hypothesis. Their inclusion here allowed for the possibility of extending some of the previous findings, particularly by the use of neurotic groups where only normal subjects had been used before. In addition, it provided an opportunity of making comparisons with the performance of psychotics on the same tests. Further tests of a broadly perceptual nature were added, and the problem of attention considered by the inclusion of a test of the vigilance type, the nature of which appeared particularly applicable to the problem of inhibition. Apart from the light which, it was anticipated, would be thrown on Eysenck's theory by each test taken individually, the application of a number of measures to the same population of subjects was, of course, expected to give valuable information about the generality of the inhibition hypothesis as an explanatory concept for extraversion.

The aims, then, of the investigation reported can be summarized as follows: (1) To investigate, on a number of different measures, Eysenck's hypothesis that extraversion and central inhibitory growth are positively related. (2) To estimate the generality of this hypothesis. (3) To study the performance of schizophrenics relative to that of other groups of non-psychotic subjects and, consequently, to suggest, if possible, how such findings may be incorporated within the general framework of Eysenck's theory.

SELECTION AND DESCRIPTION OF SUBJECTS

Four groups of subjects were included in the investigation. These were hysterics, dysthymics, schizophrenics, and normals. All the groups were selected from an Army population resident in a military psychiatric hospital.[1] In the case of the two neurotic groups, all consecutive admissions were considered for testing. They were accepted (*a*) if, in the case-notes accompanying the patient, there was consistent agreement regarding diagnosis, and (*b*) if there was agreement, consistent with (*a*), among the psychiatrists within the hospital who had seen the patient.

The hysteric group consisted of individuals whose principal condition included marked hysterical conversion symptoms (five cases), fugue states or simple amnesia (eight cases), or periodic blackouts known to be of hysterical origin (three cases).

The dysthymic group was made up of individuals who complained variously of frequent headaches, gastric disturbance, and irrational fears, and who showed marked anxiety symptoms and signs of underlying tension. No obsessionals were included in the dysthymic group.

Similar criteria were adopted in the selection of the schizophrenic group, except that an additional condition of acceptance for testing was the degree of co-operativeness of the patient. Subjects were rejected who were deemed insufficiently co-operative to be likely to complete the rather lengthy test battery involved. No particular specification was made regarding the type of schizophrenic reaction, but three patients were diagnosed as paranoid, five as simple, eight as hebephrenic, none as catatonic.

---

[1] The writer is indebted to Colonel J. McGhie, Commanding Officer at the Royal Victoria Hospital, Netley, for his kind permission to test both patient and normal subjects under his charge. The research was carried out under U.S. Army Research Grant No. D4-91-591-EUC-1136, and the writer wishes to acknowledge his indebtedness to the grant-giving body.

All but two of the normal group were male nursing orderlies in the hospital wards or technicians in the pathological laboratory. Two subjects were non-psychiatric patients from the surgical ward convalescing from minor operations.

There were sixteen subjects in each group and, apart from age, an estimate of intelligence, in terms of the Army SSG rating,[1] was available for each subject. Mean ages and mean SSG ratings for the four groups are shown, respectively, in Tables 2.1 and 2.2. Analyses of variance revealed that there were no significant differences between the groups in either age or intelligence level.

TABLE 2.1

Age

	Mean	S.D.		Mean	S.D.
Hysterics	22·84	7·036	Schizophrenics	23·38	4·609
Dysthymics	26·06	6·957	Normals	21·34	2·469

TABLE 2.2

SSG rating

	Mean	S.D.		Mean	S.D.
Hysterics	3·00	1·172	Schizophrenics	3·38	1·525
Dysthymics	2·88	0·977	Normals	2·38	1·042

Sometime during the testing session the E and N scales of the Maudsley Personality Inventory (MPI) were administered to each subject, who was also asked to complete four scales (Hy, Pt, Pd, Sc) of the MMPI. The respective means for the four groups on each scale of these two questionnaires are given in Tables 2.3 and 2.4.

A brief comment should, perhaps, be inserted here with regard to the MPI scores of the various groups, particularly the two neurotic groups. It can be seen from the mean scores for the latter groups that the dysthymics are, as expected, excessively introverted as well as being, of course, highly neurotic. There is some discrepancy in the hysteric group,

---

[1] This estimate of intelligence gives a rating on a six-point scale: 1, 2, 3P, 3M, 4, and 5, the latter being the lowest rating. For the purpose of calculation the ratings 3P and 3M were called 3 and 4, the next two ratings thus becoming 5 and 6.

TABLE 2.3

MPI

	E-Scale		N-Scale	
	Mean	S.D.	Mean	S.D.
Dysthymics	12·19	4·537	38·00	5·798
Hysterics	24·00	6·595	32·62	11·115
Schizophrenics	25·75	9·930	31·56	12·420
Normals	26·19	7·366	20·88	10·375

TABLE 2.4

MMPI

	Hy		Pt		Pd		Sc	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Dysthymics	66·75	10·485	74·06	16·231	63·06	18·564	72·50	24·132
Hysterics	62·25	14·089	65·88	17·148	63·62	12·791	71·12	17·469
Schizophrenics	60·31	13·937	72·38	15·815	65·75	14·092	76·69	23·480
Normals	51·69	9·698	57·18	13·404	53·12	11·336	54·50	13·883

however, which emerges with a rather lower mean E-scale score than would be anticipated on the basis of Eysenck's original predictions (1953b) with regard to neurosis. Hysterics appear, in fact, somewhat *less* extraverted than the normal subjects. That this discrepancy is not without a precedent is evidenced in the work of Sigal, Star, and Franks (1958a), who report similar results with an unselected sample of neurotics. The detailed implications of their findings, to be found in a discussion between the authors (1958b) and Eysenck (1958b), do not concern us here. We may simply note that, as far as the initial comparisons between the two neurotic groups are concerned, it is necessary only that the hysterics be more extraverted relative to the dysthymics, which in fact they are in this case to a degree which is significant at a high level of confidence. This is not meant to imply that further consideration of this question may not be necessary when the overall relationship between inhibition and extraversion is evaluated, particularly with regard to the outcome of predictions about the relative performance levels of hysterics and the slightly more extraverted normal group. Indeed, the results of such predictions may be expected to throw some light on the role played, if any, by variables other than inhibition.

A DETAILED CONSIDERATION OF THE TASKS USED

In addition to the two questionnaires, nine[1] objective behaviour tests were administered to all subjects. Six of the tests were chosen as being

[1] The test battery consisted originally of eleven tests, two of which were not completed by all subjects and were therefore not included in the final analysis.

directly relevant to the investigation of the inhibition hypothesis; that is to say, individual differences in performance on them could, it was predicted, be explained primarily in terms of differential rates of inhibitory growth. These six tests were named: Pursuit-rotor, Auditory Flutter Fusion (AFF), Archimedes Spiral, Time Error, Vigilance, and Time Judgement. The three other tests were less obviously related to this hypothesis, but provided objective criteria of neuroticism and psychoticism additional to those derived from the questionnaire measures. These were named: Body Sway, Sorting Test, and Picture Recognition Test. Each of these tests is now considered in more detail.

Vigilance. Those tests which may be broadly subsumed under the heading of 'vigilance' seem, for a number of reasons, to be particularly suitable for the direct investigation of the theoretical position taken here. In their simplest form, tests of this type require the subject to attend over a relatively long period to a monotonous stimulus pattern and to watch out for and respond to 'signals' occurring irregularly and at infrequent intervals. One of the main advantages, therefore, of this type of task is that a relatively lengthy sample of behaviour is forthcoming, in which performance is largely uncomplicated by the habit factor, since the simple responses involved are usually at an asymptotic level.

This approach to the problem of attention was taken originally by Mackworth (1950), using a clock-watching task. His demonstration that decrement occurs in vigilance during prolonged observation fits well an interpretation in terms of inhibition. Since then a variety of vigilance situations has been investigated and the relevant studies have recently been reviewed by Broadbent (1958). There appears to be general agreement among most studies that a decline in efficiency, reflected in a failure to detect signals, occurs over time, although a number of experimental conditions, which do not concern us here, are known to diminish or even prevent this decline. In general, it may be said that the conditions under which the greatest decline in vigilance may be expected to occur over the shortest possible time would be on tasks where the signals are infrequent and irregular, where they occur within the context of similar neutral stimuli, and where external motivating stimuli are reduced to a minimum. Since a situation in which there was maximum decline in the shortest time was most suited to our purposes, a task was chosen which met these criteria most adequately. Before describing the task used, however, brief consideration must first be given to the general nature of vigilance and its relation to personality.

Space limits a detailed analysis of vigilance decrement here, but a number of points are worth noting. The assumptions underlying the predictions to be made here—viz. that decline in vigilance is due to inhibitory processes—almost certainly implies that both forms of inhibition—reactive and conditioned—are likely to play a part in determining performance. The demonstration by Bills (1931), for instance, of 'blocks' or rest pauses in continuous work, reformulated in Hullian terms, has led to an analysis (e.g. Eysenck, 1956b) which suggests that such rest pauses serve the purpose of partially dissipating reactive inhibition which has hitherto accumulated. This dissipation of negative drive (I_R) acts as a reinforcement for the rest pause and so, in the usual way, for the formation of the habit of not responding or conditioned inhibition ($_SI_R$). It is very likely that such an analysis is applicable to the problem of vigilance, where such breaks in attention clearly reduce the probability of detecting signals. However, whatever the relative importance of these two variables, the general predictions about individual differences in vigilance remain unaffected, since a more rapid growth of reactive inhibition will lead, other things such as positive drive being equal, to an earlier onset, and subsequently greater frequency, of rest pauses. Accordingly, hysterics should show more decrement than dysthymics, and normal extraverts more decrement than normal introverts.

There is available some evidence confirming the essential correctness of this suggestion, although all of it appears to have been gathered on normal subjects. The most relevant of these is a study by Bakan (1960; pers. comm.), using a task on which the one employed here was based. He showed that in the first thirty-two minutes of an eighty-minute vigilance task introverts were superior to extraverts. This difference disappeared when a secondary task was introduced, largely because extraverts benefited very much more than introverts from this condition. The latter finding would seem, incidentally, to fit an explanation in terms of disinhibition, since the extraverts, being more likely to have high levels of inhibition at any one stage in the task, would be expected to benefit most from the disinhibiting effects of the buzzer response required to the secondary stimuli.

Other evidence supporting the viewpoint that vigilance decrement is greatest among extraverts, is found in a number of unpublished studies reported by Broadbent (1958), who also quotes his own finding that a measure of level of aspiration, assumed to be positively associated with introversion, is correlated to the extent of −0·46 with deterioration in performance on his twenty dials test. It is clear from the foregoing that it was possible to predict with a fair degree of confidence the vigilance performance of the normal and neurotic groups used here.

The prediction was made, therefore, that hysterics would reveal on the vigilance task inferior performance to that of dysthymics, the normal group falling between the two neurotic groups. It was also predicted that a negative correlation would be found between vigilance performance and extraversion in the normal group. Since, following Bakan's procedure, a secondary task condition was also included in the test, the assumption that this secondary task acted as a disinhibitor led to the further prediction that its introduction would produce a greater increase in the performance level of hysterics than of dysthymics, thus probably causing the difference between the two groups to disappear. Similar but less marked changes were also expected to occur within the normal group, extraverts improving more than introverts.

Predictions concerning the performance of the schizophrenics on this task could be made with less·certainty, although clinical impression would doubtless lead us to the view that they would be inferior to normal subjects. The consistent findings, concisely summarized by Hunt and Cofer (1944), that the reaction time of schizophrenics is abnormally slow are relevant, since the reaction time and vigilance situations clearly have some elements in common. If, as we considered earlier, the schizophrenic's reaction time is, indeed, a result of poor attention, as the work of Huston, Shakow, and Riggs (1937) and Rodnick and Shakow (1940) suggests, then it is not unreasonable to assume that the schizophrenic will perform badly on tests such as vigilance, which are essentially dependent on the ability to sustain attention. An analysis in terms of 'protective inhibition' would also lead us to essentially the same view. It seemed reasonable to propose the hypothesis, therefore, that a group of schizophrenics would perform at a significantly worse level than normals on the vigilance task.

It would perhaps be carrying the analogy between vigilance and reaction time too far to suggest that we could predict the outcome of the introduction of the secondary task in the case of the schizophrenic group, although mention should be made of the work of Tizard and Venables (1957), who have demonstrated that the reaction time of some schizophrenics can be decreased by the application of extraneous stimulation. It must be noted, however, that the authors have rejected an explanation of this in terms of disinhibition. Because of the lack of direct evidence on this point no specific prediction was made regarding the effect of the secondary task on the performance of the psychotic group.

The vigilance task used was in two parts, the first consisting of a thirty-minute tape-recording of digits read out at the rate of one per second. Some departure from randomness was unavoidable, but the series did not differ noticeably from a random one. Interspersed in the series were a number of 'signals', which consisted of three successive odd digits. The signals occurred with a frequency of one per minute, the exact position of each signal within each minute being determined randomly. The mean interval between signals was 56 seconds, with a range from 3 seconds to 1 minute 41 seconds. During this part of the test the subject was instructed to listen particularly for the groups of three odd numbers, which, he was told, might occur at any time during the test. He was told to tap loudly on the table whenever he heard such a group. He was then given a short practice series and, if necessary, the instructions were supplemented by further explanation until it was clear that the procedure had been understood.

Immediately following the principal part of the test another series of digits was played, lasting this time for ten minutes. Here the subject was told to carry on as before but was also instructed to listen out for the number 'six' in addition to the groups of three odd numbers. He was asked to press a buzzer in front of him in response to the number 'six' as well as tapping in response to the main signals. The latter were again interspersed with the same frequency as before, the mean interval between signals being 53 seconds, with a range from 8 seconds to 1 minute 40 seconds. The secondary signals occurred much more frequently, the number 'six' appearing in the series once every seven seconds.

During testing the tone and volume controls of the tape-recorder were adjusted to a standard setting for all subjects, the setting being determined beforehand so as to give a clear, audible reproduction. Throughout the test the experimenter sat in a position as unobtrusive as possible and marked off on a prepared scoring sheet all signals detected by the subject.

The two test scores taken for the initial analysis of results were the total numbers of primary signals detected in each part of the test. Totals were also taken for each five-minute period of the test.

Group means for the total score measure in each part of the test are shown in Table 2.5. Analysis of variance revealed that there was an overall significant difference between the four groups in both parts of the test, the confidence level reached in each case being 0·001.

TABLE 2.5

Total vigilance score

	First part		Second part	
	Mean	S.D.	Mean	S.D.
Normals	21·06	4·02	7·69	2·13
Dysthymics	24·75	4·32	5·69	1·83
Hysterics	15·44	6·09	5·19	2·63
Schizophrenics	12·75	11·62	4·19	4·11

The striking differences between the four groups, shown in the total mean scores, are also reflected in their respective performance levels throughout the test, as revealed in Fig. 2.1. It is clear from this figure that, as expected, dysthymics were superior at all stages in the first part of the test, while hysterics performed at an inferior level throughout, again as predicted. Psychotics also performed badly, although here there were much larger individual differences than in the other groups, as the standard deviation shows. In terms of total score during the first part of the test, dysthymics were significantly better than normals

($t = 2{\cdot}4117, p < 0{\cdot}02$), while the normals were significantly better than the hysterics ($t = 2{\cdot}9946, p < 0{\cdot}01$).

A product moment correlation calculated for the normal group between total vigilance score and the E-scale score of the MPI had a value of + 0·19, which was not significant, but in any case was in a direction opposite from that predicted. Further inspection of the data suggested that the total score measure of vigilance taken initially, while sufficiently sensitive to confirm the overall group differences, was actually a relatively crude one reflecting, as it did, two distinct aspects of the task, viz. performance level and rate of fall-off in level. In order to separate these two components linear regression lines were fitted to the

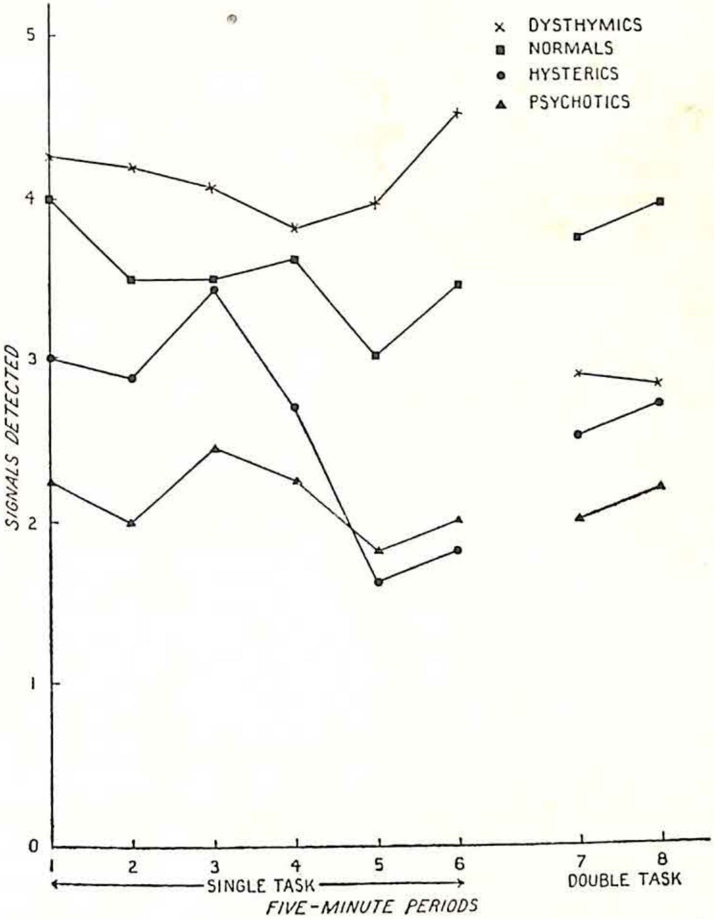


FIG. 2.1. Vigilance: Mean Performance for each Group in each five-minute Period of the Task.

data for each subject.[1] This gave two measures, the coefficient '*b*' representing performance level (strictly starting level) and the coefficient '*a*' representing mean change per trial, which, in the case of negative values of '*a*', may be interpreted as 'rate of fall-off'. The mean values for these two coefficients for the four groups are shown in Table 2·6.

TABLE 2.6

Vigilance regression coefficients

	'*a*'		'*b*'	
	Mean	S.D.	Mean	S.D.
Normals	−0·119	0·304	3·93	1·106
Dysthymics	+0·007	0·236	4·10	1·644
Hysterics	−0·297	0·353	3·61	1·621
Schizophrenics	−0·057	0·182	2·32	2·033

It may be noted in passing that both measures confirm the inter-group differences demonstrated previously. Our immediate concern being with the normal group only, however, correlations were calculated between the E-scale score and the two coefficients for this group. The correlation between fall-off ('*a*') and E score was +0·45 and between E score and level ('*b*') −0·31. Both of these just failed to reach significance at the 0·05 level for sixteen subjects, but the direction of the correlations was suggestive that there was a tendency for introverts to start at a higher level than extraverts but to fall off more. The situation is more clearly shown in Fig. 2.2., where the performance curves for the six most introverted and six most extraverted normals are compared with those of the two neurotic groups. It can be seen from this that hysterics and normal extraverts and dysthymics and normal introverts showed the expected similarity only in so far as they started at comparable levels. It may be anticipated that this finding is of some significance, but a discussion of it is delayed until a later section, when a general interpretation of the results is presented.

The changes consequent upon the introduction of the secondary task are shown clearly in Fig. 2.1 and in Table 2.5. The hysteric group showed a significant *increase* in performance level, t being 2·9213, $p < 0·02$, while the dysthymic group showed a significant *decrease* in level, t equalling 9·3258, $p < 0·001$. The results of this change were that the previous significant difference between the two neurotic groups disappeared, as predicted, t being 0·609 (NS). The change in the normal group was significant at the 0·1 level only, t being 1·9211, and the relationship of this change to extraversion was opposite to expectation but not significant, r equalling −0·17. The psychotic group showed no change in performance level.

---

[1] For this purpose the data as exemplified in Fig. 2.1 were used, i.e. there were six 'trials', each representing a five-minute period.

In general, then, it may be said that the vigilance test differentiated between normal, neurotic, and psychotic subjects in a predictable manner, thus lending overall support to the inhibition hypothesis which led to its use.

Pursuit rotor. A further approach to the problem of inhibition, mainly reactive inhibition, is found in work with continuous motor tasks such as the pursuit rotor. The assumption that the phenomenon of reminiscence, which can readily be demonstrated on tasks of this kind, is directly due to the dissipation of reactive inhibition has led to a deduction, which Eysenck (1956b) has confirmed with normals, that extraverts, accumulating more inhibition, should show more reminiscence following a rest than will introverts. Following this finding the prediction was made, therefore, that in the present study hysterics would show greater amounts of reminiscence than dysthymics, with normals intermediate. Within the normal group reminiscence and extraversion were expected to be positively correlated.

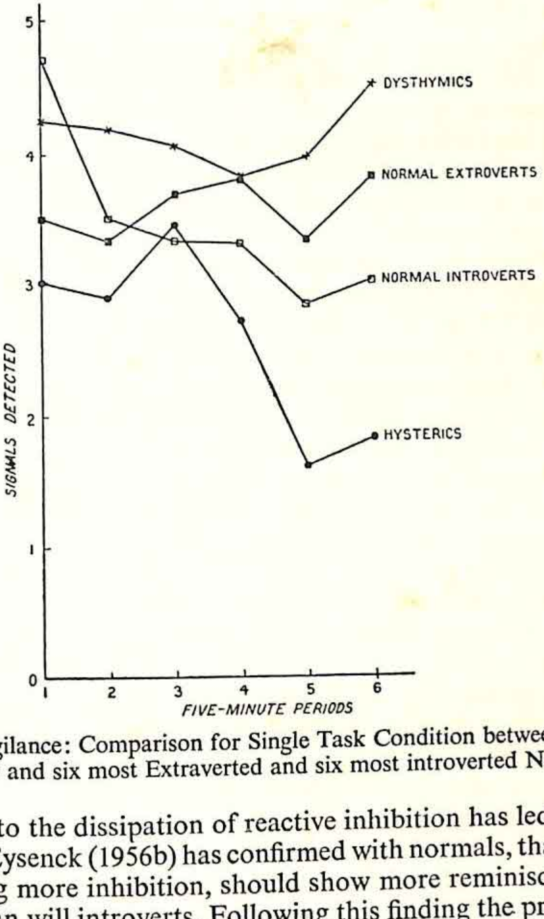


FIG. 2.2. Vigilance: Comparison for Single Task Condition between Neurotic Groups and six most Extraverted and six most introverted Normals.

As with the vigilance test, the performance of schizophrenics, particularly in respect of reminiscence, was difficult to anticipate. The expectation that their general level of performance would be lowered relative to that of normal subjects was certainly sustained by the study of pursuit rotor learning in psychotics by Huston and Shakow (1948), who showed that schizophrenics were very inferior on a task similar to the one used here. These authors did not, however, specifically investigate reminiscence, although they did find incidentally that retesting after an interval of three months resulted, in the schizophrenic group, in a slight increase analogous to a reminiscence effect; normals, by comparison starting the second session lower than the level they had reached at the end of the first. It is difficult to estimate the significance of these findings and, in any case, the testing conditions are not sufficiently comparable to those obtaining here to provide the basis for a definite prediction about reminiscence in the psychotic group. The most appropriate hypothesis in this instance seemed, therefore, to be the null one that schizophrenics and normals would not differ essentially in reminiscence.

The pursuit rotor apparatus is too well known to require detailed description, and a full account of the instrument used and instructions given is contained in an unpublished thesis by Star (1957). Briefly, it consists of a rotating bakelite turntable, inserted near the periphery of which is a small metal disc which rotates with the turntable at a speed of one revolution per second. The subject is instructed to keep the tip of an articulated rod in contact with the small disc as it rotates in front of him. Contact between the disc and the metal tip of the rod activates a recording apparatus consisting of two electric clocks, which come into circuit successively every ten seconds. This enables a continuous record to be taken of the amount of time during each ten seconds that the subject is on target. The measure is normally, as here, converted to a measure of percentage time on target.

The schedule followed here consisted of three five-minute practice periods each being separated by a ten-minute rest pause. During the rest pauses all subjects were asked to complete the two questionnaires used in the investigation. In addition to the individual scores for each ten-second period of the three phases a reminiscence score was calculated for each subject. This was done by subtracting the mean of the last three scores before the rest period from the first score after the rest period. Two such reminiscence measures were, of course, taken.

Fig. 2.3 shows the performance curves of the four groups in the three phases of the test. Analysis of variance carried out for the three phases indicated that there was an overall significant difference ($p < 0{\cdot}001$) in performance between the four groups. Numerically, the performance measures for the groups are most concisely represented in the coefficients for linear regression, which were ultimately computed on the data for each subject in each phase of the test. The group mean values for these coefficients are shown in Table 2.7. The aspects of the performance curves which we may note now, but which will concern us in more

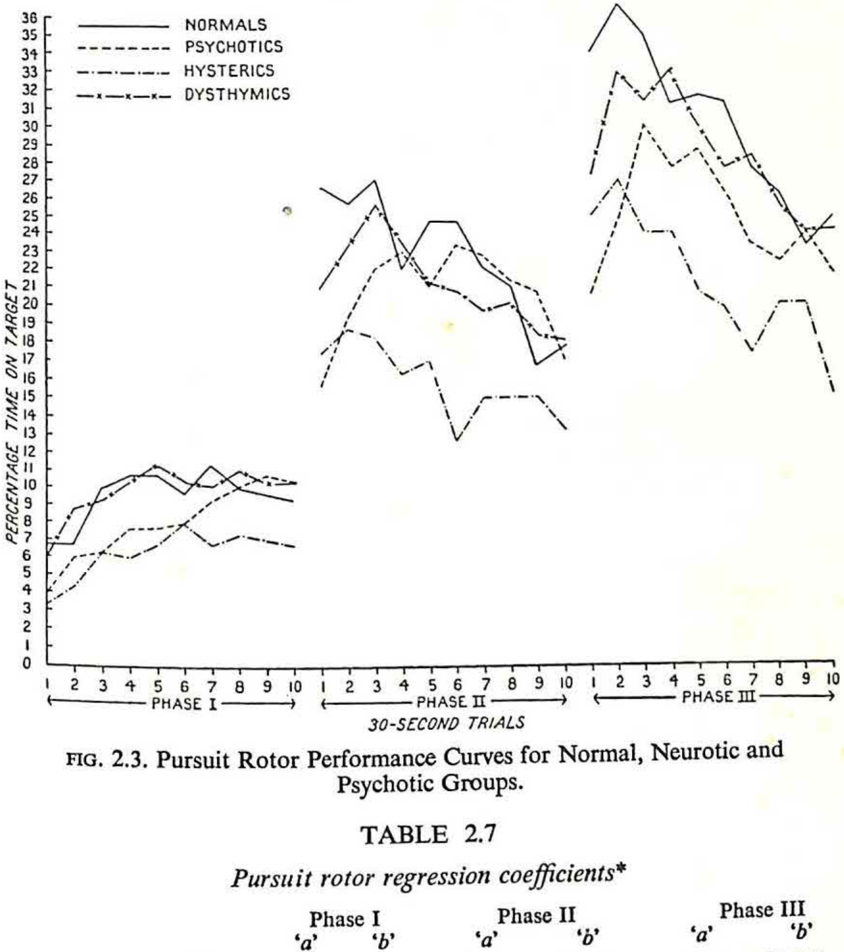


FIG. 2.3. Pursuit Rotor Performance Curves for Normal, Neurotic and Psychotic Groups.

TABLE 2.7

*Pursuit rotor regression coefficients**

		Phase I		Phase II		Phase III	
		'*a*'	'*b*'	'*a*'	'*b*'	'*a*'	'*b*'
Normals	Mean	0·2	7·88	−0·87	27·80	−1·38	37·86
	S.D.	0·436	7·205	0·982	9·274	0·815	12·335
Dysthymics	Mean	0·34	7·91	−0·63	24·66	−0·81	32·90
	S.D.	0·541	6·670	0·995	12·510	0·912	14·066
Hysterics	Mean	0·35	4·18	−0·50	18·66	−1·07	27·19
	S.D.	0·342	5·699	0·601	12·193	0·969	16·466
Schizophrenics	Mean	0·67	3·68	0·13	19·89	−0·15	25·55
	S.D.	0·663	6·508	0·767	13·946	0·817	14·948

* '*a*' refers to 'slope' or 'average increment per trial' and '*b*' to 'performance level'.

detail later, are the low and high performance levels of hysterics and dysthymics respectively, and the rather unusual performance curves of the psychotic group, particularly during the first phase of the test. Our main predictions, however, were made with regard to the reminiscence scores, known respectively as RS1 and RS2. The group mean values for these two measures are shown in Table 2.8. While the groups were found on analysis of variance to be significantly differentiated at the 0·001 level, the direction of the group differences was, as can be seen from Table 2.8, somewhat different from that predicted. Although dysthymics followed the prediction that they would show less reminiscence than normals, hysterics failed to confirm our expectation that they would have higher reminiscence scores than normal. Dysthymics and hysterics were virtually identical on the RS1 measure, while the difference in RS2, although in the right direction, is not significant. Within the normal group the correlations between reminiscence and extraversion were not significant, the values for *r* between RS1 and RS2 and the E scale being, respectively, −0·12 and −0·09. One notable feature, and the source of the overall significant difference, is the low degree of reminiscence shown by the schizophrenic group, the differences between the latter group and the normal group being significant for both RS1 and RS2 (*t* equals 4·0689, $p < 0·001$ and 2·4222, $p < 0·05$, respectively). While not anticipated, this result appears to merit further discussion which will be taken up later when other, essentially similar, findings, obtained contemporaneously with our own, will be presented.

TABLE 2.8

Pursuit rotor reminiscence scores

	RS1		RS2	
	Mean	S.D.	Mean	S.D.
Normals	13·7	8·51	14·8	11·53
Dysthymics	8·3	7·91	6·2	14·07
Hysterics	8·0	9·76	12·3	9·67
Schizophrenics	1·9	7·62	3·9	13·53

Archimedes Spiral. The applications, on the one hand, of the inhibition hypothesis to the explanation of the learning process and, on the other, of the satiation hypothesis to perceptual phenomena, such as figural after-effects, have until recently tended to proceed somewhat independently of each other. There is, however, increasing agreement that inhibition and satiation are essentially similar in effect and due to the same underlying process (Duncan, 1956), in so far as satiation, like inhibition, may be thought to arise centrally as a result of afferent stimulation and to cause a temporary resistance to the continuation of such stimulation. One method of proof of this hypothesis is the differential one suggested by Eysenck (1957) whereby, either correlations are demonstrated between measures of inhibition and measures of satiation gathered on the same population, or measures of each are shown to correlate with some mediating variable such as extraversion, which, because of its positive relationship with inhibition, should also correlate with the degree of satiation shown on perceptual tasks. Proof of the latter kind is exemplified by Eysenck's demonstration (1955a) that introverted neurotics show less kinaesthetic figural after-effect than extraverted neurotics. Differential proof of both kinds would be expected to be forthcoming from the present study where a number of measures relevant to both the inhibition and satiation hypotheses were taken on the same groups of subjects.

The particular phenomenon known as the Archimedes spiral after-effect falls into a class of illusions in which prolonged observation of a moving stimulus, in this case a rotating spiral, results in a marked after-image when the stimulus is terminated. This after-image takes the form of apparent movement in the direction opposite to that obtaining during the original stimulation. One explanation of this in terms of satiation, suggested by Eysenck (1957), is that 'the original stimulation sets up certain unspecified cortico-neural events which are perceived as the illusion. According to satiation theory, these cortico-neural events must produce inhibition in the structures mediating the effects, so that eventually the after-image is brought to a stop.' A more consistent explanation, possibly, is that the original stimulation itself produces inhibition in addition to the excitatory state set up simultaneously. The after-effect being, it is assumed, dependent on the strength of excitation at the point when the stimulation is terminated, the length of the subsequent after-image would then be determined by the balance, at that point, between the excitatory and the opposing inhibitory potentials. Whichever explanation is correct, the prediction would be the same, that the length of the after-effect found would be negatively correlated with extraversion. Work with this test on the brain-injured, notably that of Price and Deabler (1955) and Gallese (1956), and to some extent that of Holland and Beech (1958), tends to confirm this, in so far as these subjects showed diminished after-effects consistent with the extraverting effects (Petrie, 1952) of brain damage. The authors of the first of these two papers also agree that functional psychotics show after-effects which do not differ from normal.

It was predicted here that hysterics would show significantly shorter after-effects than dysthymics and that, consistent with this, there would be negative correlations within the normal group between length of spiral after-effect and extraversion. It was further hypothesized that schizophrenics would not differ essentially from normals in the length of their after-effect.

The apparatus used consisted of a single-throw 180 degree spiral mounted on a spindle driven by an electric motor, the speed of which could be accurately controlled by means of a strobe attachment. A reversal switch allowed the spiral to be rotated in both a clockwise and an anticlockwise direction. The test was administered in a partially darkened room and the spiral illuminated by a 100-watt bulb suspended about nine inches above it. Strict control of illumination was difficult to obtain and this, and other factors such as angle of regard, were not rigidly controlled, but, in any case, variables other than the length of stimulation have been shown (Holland, 1958) to have a minimal influence on the spiral after-effect. The subject was seated about four feet in front of the apparatus and asked to fixate the centre of the spiral while it was rotating. He was told to continue fixating after the spiral was stopped by the experimenter, and to report immediately the after-effects of it ceased. The length of the after-image was measured with a stop-watch. Four trials were given in all, alternately clockwise and then anticlockwise. In each trial the spiral was rotated for one minute and a rest of one minute was given between trials. Four scores were, therefore, obtained from each subject.

Analysis of variance indicated, in addition to a significant overall difference between groups ($p < 0·001$) that there was no difference in after-effect score due to direction of rotation. The mean of four trials was, therefore, taken as the measure of performance on this test, the group means for this measure being shown in Table 2.9.

TABLE 2.9

Spiral after-effect

	Mean	S.D.
Normals	10·16	4·560
Dysthymics	15·78	5·486
Hysterics	9·74	5·337
Schizophrenics	14·78	6·996

As predicted, hysterics showed a shorter after-effect than dysthymics, the difference being significant (t equals 3·0659, $p < 0·01$). Dysthymics had a longer after-effect than normals, again as predicted, t being equal to 3·0504, $p < 0·01$. Hysterics did not differ markedly from normal in length of after-effect, although, perhaps surprisingly in view of the previous results quoted, the test differentiated between the normal and schizophrenic groups (t equals 2·1488, $p < 0·05$), the latter having a longer after-effect than normal. Again it must be noted in this respect that, as with most other measures taken, the standard deviation for the psychotic group was somewhat larger than that for other groups, indicating a lack of consistency in the behaviour of schizophrenics as a group. This is not surprising in view of the wide variety of clinical reactions subsumed under this heading.

Within the normal group the length of after-effect was found to be significantly related to extraversion, r between the E score and length of after-effect being −0·596, which for sixteen subjects is significant at the 0·01 level. This finding, together with the inter-group differences outlined above, thus provides fairly substantial support for most of the hypotheses proposed with regard to this test.

Time Error. The presence of a particular constant error, known as time error, in judgements of sucessively presented stimuli has concerned psychophysicists for many years. Briefly, this error takes the form of a systematic tendency to overestimate or underestimate the magnitude of the second of two physically identical stimuli occurring successively. The effect has been investigated, although not always demonstrated, using a variety of stimulus material, including the pitch and intensity of sound (Postman, 1946), light intensity (Kreezer, 1938), hefted weights (Guilford and Park, 1931), and visual extent (McClelland, 1943). In general, though not universally, the findings have been that (*a*) negative time errors—the tendency to overestimate the second of a pair of stimuli—tend to predominate over positive ones; and (*b*) the tendency towards negative time error increases as the interval between the two compared stimuli increases (Kohler, 1923; Needham, 1935).

A number of theoretical explanations, none entirely satisfactory, have been proposed for this constant error and these have been adequately reviewed by Osgood (1953). The investigation of the problem here derives from its relevance to the satiation, and therefore the inhibition, hypothesis, following Kohler's original trace theory of time error. An analysis along these lines would be formulated as follows. The application of the first of a pair of stimuli, say a tone of a certain intensity, will produce, in addition to the expected excitatory effect, an inhibitory process opposing the continuation of this excitation. A following tone of equal intensity will then be judged louder according to the degree to which the inhibition consequent upon the first tone has 'damped down' the original excitation. One of the factors determining this would be, as the evidence confirms, the duration of time before the second stimulus is presented or, as it is usually called, the S–V interval. One difficulty with this general explanation of time error as it stands is the demonstration in some studies that where the S–V interval is very short, say up to two seconds, positive or zero errors may tend to replace the more usual negative errors. This is so in the case of judgements of sound intensity (Postman, 1946), with which we are particularly concerned here. A possible explanation of this is that, when the second stimulus follows the first very rapidly, the original inhibition induced tends to oppose what is effectively a continuation of the first stimulus. In order to be judged *equal*, in fact, the second stimulus would then have to be of a physically greater intensity, thus giving rise to positive time error.

In addition to the general influence of the S–V interval on the degree of time error it would be expected, if our argument so far is correct, that individual differences would appear and that these would be related to extraversion. Extraverts, producing more inhibition as a result of the first stimulus, would be expected to show a greater tendency than introverts to overestimate the intensity of the second of two stimuli. Accordingly the general prediction was made here that hysterics would show a greater tendency towards negative time error than dysthymics, while normals would tend as a whole to fall between these two groups of subjects. It was also predicted that extraversion and tendency towards negative time error in normals would be positively correlated. The appearance at the shorter S–V intervals of positive time error in some subjects was not unlikely, although exact predictions about group differences in this respect could not confidently be made. It could, however, be predicted that there would be a general tendency in all groups for negative time error to increase as the S–V interval increased, and it was expected that this change would be greatest in hysterics and least in dysthymics.

No specific prediction was made with regard to the schizophrenic group, although the 'protective inhibition' approach to schizophrenia would lead us to expect larger negative time errors in these subjects than in normals particularly if, as was the case here, the stimuli involved are relatively intense.

The stimulus material used in the investigation of time error consisted of a tape-recording of pure tones, all of a frequency of 1100 c.p.s. The material was divided into four series corresponding to the four S–V intervals of 1, 2, 4, and 6 seconds. The standard tone, which was always presented first, had an intensity of 75 decibels and a duration of 1 second. Within each series there were five variable tones, all of one second duration and with respective intensities of 73, 74, 75, 76, and 77 decibels. Each of these was paired ten times with the standard in each series, the order of presentation of pairs being determined randomly. There were thus fifty judgements to be made in each of the four series, with an interval of six seconds between pairs and a rest of five minutes between series. The stimuli were presented to the subject through head-phones and the volume of the tape-recorder was adjusted to the correct setting by means of a calibrated meter.

The subject was instructed to decide as quickly as possible whether the second sound of each pair was louder or softer than the first sound, no 'equal' judgements being allowed. He was then asked to encircle the appropriate response on the answer sheet. The measure of time error taken was calculated according to the following formula:

$$\frac{(\text{Number of judgements 'softer'}-\text{number of judgements 'louder'}) \times 100}{\text{Total number of judgements}}$$

There were, of course, four such measures for each subject.

The average degree and direction of time error shown by the four groups at each S–V interval can be seen in Fig. 2.4. While, as expected, there is in all subjects a general tendency towards negative, rather than positive, time error, this tendency can be seen to be greatest among

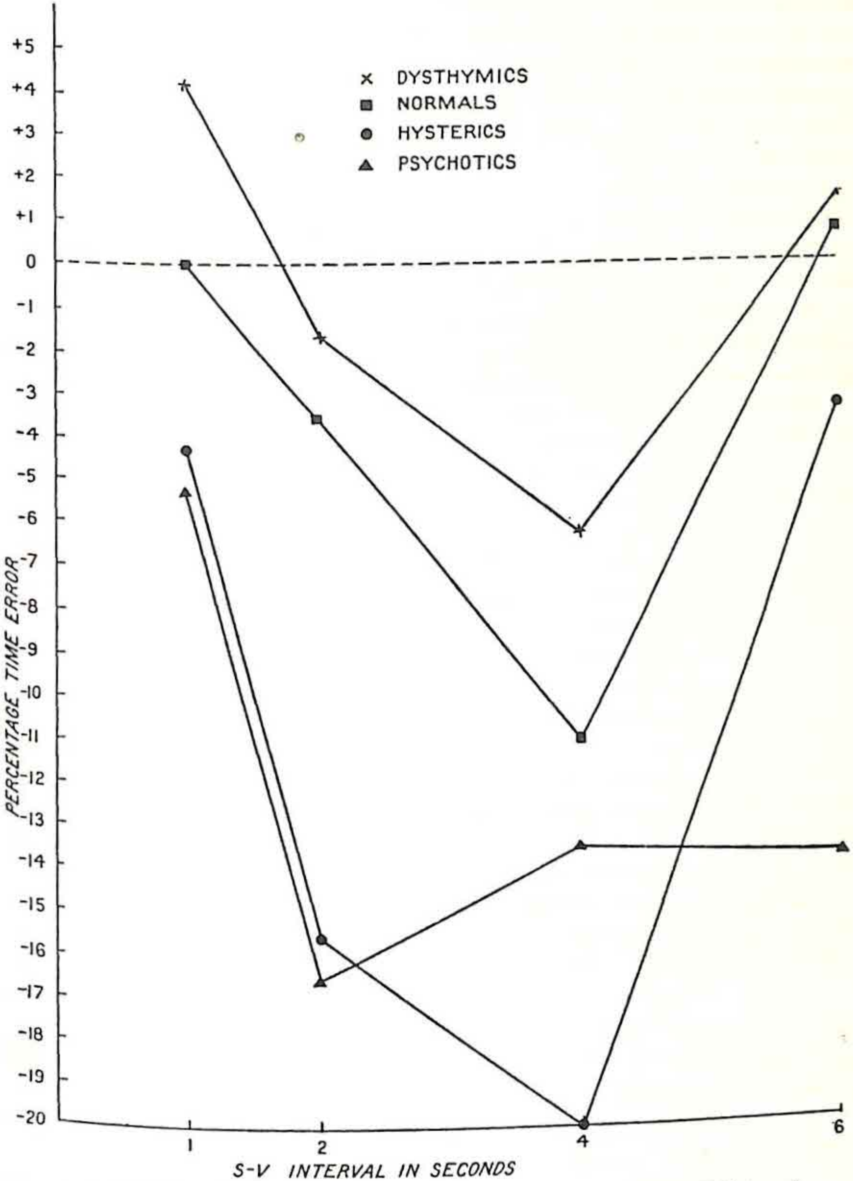


FIG. 2.4. Time Error—Mean Error for each Group at four S–V Intervals.

hysterics and psychotics and least among dysthymics, with normals intermediate. It can also be seen that, confirming our predictions and previous work, the time error effect becomes more negative the longer the S–V interval and this occurs to a differential extent among the four groups, which tend to diverge. The qualification to this statement is, of course, that at the six-second S–V interval all groups, except the psychotics, reveal a movement *away* from further negative time error and tend to approach a level of zero error. The explanation of this is unclear, although one possibility is that, at this and longer time intervals, the interaction between excitatory and inhibitory processes, which we have proposed to underly this phenomenon, becomes minimal and judgements tend to be made on a random basis. An interesting possibility, which might support this, is the differential reaction shown by the psychotic group. If, as we consider earlier, 'protective inhibition' underlies the schizophrenic's performance on this test and if, as has recently been suggested by Venables (1959), it is the rate of dissipation of this inhibition which is critical in schizophrenia, then it would not reach a minimal level until an S–V interval is reached which is longer than that required to show the zero effect in normals. The meaningfulness of such an interpretation must, however, await further experimentation.

TABLE 2.10

% Time error measure

	Mean	S.D.
Normals	−3·44	8·661
Dysthymics	−0·56	10·868
Hysterics	−10·78	11·306
Schizophrenics	−12·25	15·237

A two-way analysis of variance revealed that the four groups were significantly differentiated at a confidence level of 0·001. A single measure of time error was then obtained by taking the mean of the values for the four S–V intervals. The group means for this measure are shown in Table 2.10. A number of calculations based on this measure revealed that dysthymics and hysterics were significantly differentiated (t equals 2·5297, $p < 0{\cdot}02$), as were normals and hysterics (t equals 2·0618, $p < 0{\cdot}05$). Normals and psychotics were differentiated at the 0·1 level only (t equals 2·0109) while normals and dysthymics did not differ significantly in mean time error. The correlation between the time error measure and E score for the normal group had a value of −0·24, which, while not significant, indicated, as predicted, some tendency for the more extraverted in this group to show greater negative time error.

Time Judgement. All of the tasks considered so far involve the measurement of some kind of performance change due, it is hypothesized, to the temporal growth of inhibition consequent upon particular kinds of positive test stimulation. Measures of a rather different sort, however, might be expected to give an indication of the state of the subject with respect to the inhibitory process. If, for instance, we take a task in which the subject is asked to reproduce unfilled time intervals it might be expected that satiation arising from the limited stimulation available will be one of the factors determining his degree of under- or over-reproduction of the interval. In mentalistic terms the subject will become more or less rapidly bored, the rapidity with which this occurs being determined, we would suppose, by the position of the subject on the extraversion-introversion continuum. What might be thought to be such behavioural concomitants of an underlying inhibitory process has certainly been associated in a rather different situation with extraversion (Wyatt and Langdon, 1937). Applying this to the present problem, it seems possible that the individual differences which have been noted in time perception (Woodrow, 1933) may be meaningfully analysed within the present theoretical framework.

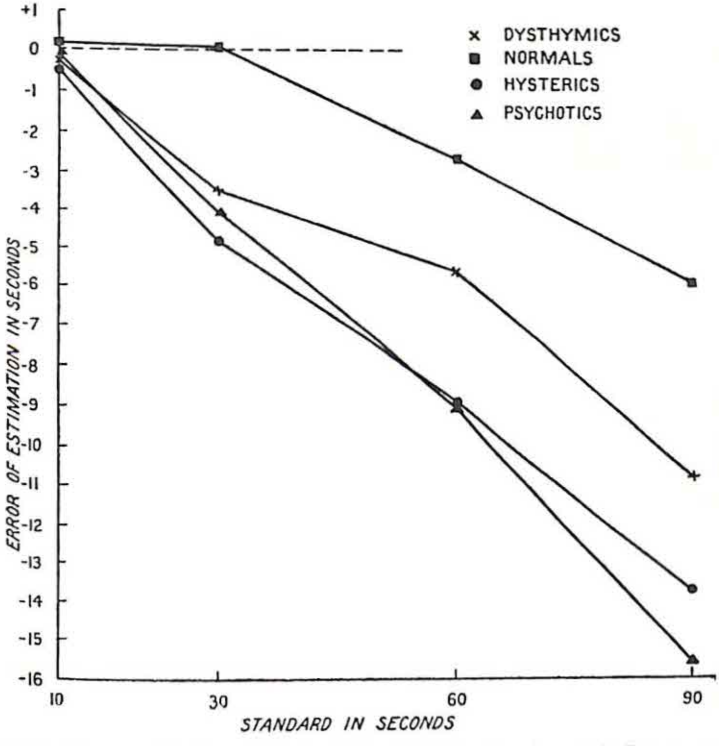


FIG. 2.5. Time Judgement—Mean Error of Estimation for each Group at four Time Intervals.

In order to investigate this a simple test was designed in which the subject was asked to reproduce a number of time intervals of varying duration. These were 10, 30, 60, and 90 seconds, each being presented twice. The beginning and end of each time interval was defined by the experimenter tapping on the table with a pencil. The subject was instructed to tap a third time when he thought the same amount of time had passed following the experimenter's second tap as had elapsed between the two taps. The time allowed to elapse before the subject's response was measured with a stop-watch and the degree of error calculated by subtracting the actual from the estimated time. A negative value thus indicated under-reproduction and *vice versa*. The mean of the two values for each of the four time intervals was taken as the measure of response for that interval. During testing the subject was separated from the experimenter by a low screen and his watch removed. He was asked specifically not to count or try to 'measure' the time actively, but simply to try and imagine what the intervals were.

It was predicted that hysterics would show a greater tendency towards under-reproduction of the intervals than would dysthymics and that the more extraverted normals would behave more like hysterics, while the introverted normals would behave like dysthymics. No specific prediction was made regarding the schizophrenic group.

The overall differences between the groups on this test were not significant, the F-ratio just failing significance at the 0·05 level. Results were, however, roughly in the direction predicted, as shown in Fig. 2.5. Hysterics showed greater tendency to underestimate the time intervals at all levels than did dysthymics, although the latter, contrary to expectation, tended to show more error than the normal group. The group means for a single measure of time judgement—the mean of four time interval scores—are shown in Table 2.11.

TABLE 2.11

Time judgement measure

	Mean	S.D.
Normals	−2·11	6·641
Dysthymics	−5·07	5·157
Hysterics	−7·01	6·931
Schizophrenics	−6·38	12·713

Within the normal group correlations indicated that, with the exception of the 90-second time interval, extraverts tended to underestimate the intervals to a greater extent than did introverts. The correlations were as follows:

'r' between E-scale and error

10-second interval: −0·51 *p* 0·05 (N = 16)
30-second interval: −0·66 *p* 0·01 (N = 16)
60-second interval: −0·51 *p* 0·05 (N = 16)
90-second interval: −0·01 n.s. (N = 16)

The correlation between the overall measure and E-scale had a value of −0·497 which is significant at the 0·05 level.

Auditory flutter fusion. The relevance of auditory flutter fusion to the theoretical position taken here arises essentially from the assumption that individual differences in the perception threshold for auditory flutter are explicable in terms of differential rates of growth and dissipation of the inhibition arising as a result of stimulation by a sound signal. The exact determinants of this threshold have been suggested in the work of Miller (1948), Miller and Taylor (1948), and Symmes, *et al.* (1955). Their analysis and its relevance to Eysenck's theory have been discussed in detail by Holland elsewhere in this book, and its repetition here will be confined to a brief outline of the main theoretical points. It has been suggested by the above authors that in order for flutter to be discriminated from fusion in interrupted noise the 'off' period of the sound/time fraction must be sufficiently long to allow the auditory sensation to decay one detectable step or j.n.d. Thus, if signal intensity is constant, for any value of the sound/time fraction the interruption rate of a burst of noise at which flutter is observed (i.e. the flutter threshold) will depend entirely on the rate of decay of the auditory sensation. Individual differences in decay rate should, therefore, be paralleled in differential flutter thresholds. Since it is argued here that the rate of decay of the auditory sensation during the 'off' period of the sound/time fraction is a function of disinhibition, then extraverts, dissipating inhibition slowly, would be expected to have lower flutter-fusion thresholds than introverts, who dissipate inhibition quickly.

In the present study it was, therefore, predicted that hysterics would show significantly lower flutter-fusion thresholds than dysthymics, with normals falling in an intermediate position. It was also expected that there would be negative correlations within the normal group between extraversion and the threshold value.

The test used was the University of Chicago Sound Discrimination Test, details of which are contained in the manual accompanying the test. Briefly, it consists of a recorded tape of interrupted 'white noise' having square wave characteristics. The tape was played to the subject through earphones at an intensity of 70 decibels, the volume of the tape-recorder being adjusted to this intensity with the aid of a calibrated meter. The repetition rates used varied from 10 to 150 i.p.s. in steps of 10 i.p.s., and the sound/time fraction had a constant value of 0·90. Four series were given, within each of which the fifteen frequencies were paired at random with bursts of continuous noise. The burst duration for continuous and interrupted noise was two seconds and the intra-pair delay one second. The subject was simply asked to indicate after each pair whether the two sounds were the same or different. The threshold for flutter fusion which was derived from the total responses constituted the subject's score on the test.

The results for this test are shown in Table 2.12. Contrary to expectation, the differences between the four groups were not significant on analysis of variance. Indeed, the difference between the hysteric and dysthymic groups was negligible, both having lower AFF scores than the normal group. The correlation between E score and AFF threshold in the normal group was also insignificant, r being −0·09. The most noticeable feature of the results is that the test seems to discriminate not so much between introverts and extraverts as between the normal and psychiatric groups. This seems to parallel findings for visual perception (e.g. Krugman, 1947) which have indicated that the CFF is also lowered in psychiatric patients compared to normal subjects. While we shall have occasion to refer briefly to this point again, it must, of course, be concluded here that none of our predictions relating to the AFF test was confirmed. Consequently, it may be said that there is no evidence from the present results, at least, that individual differences in flutter threshold can be related in a predictable manner to extraversion, although findings more in line with prediction are reported elsewhere in this book by Holland.

TABLE 2.12

AFF score

	Mean	S.D.
Normals	76·56	23·241
Dysthymics	65·31	20·354
Hysterics	65·00	15·811
Schizophrenics	60·31	21·691

Picture Recognition Test. This is a test devised by Brengelmann (1958), whose application of it to normal, neurotic, and psychotic subjects indicated that the overall difference in the scores of these groups was highly significant, normals being clearly differentiated from the two patient groups. The latter did not differ markedly from each other and a simultaneous test of Eysenck's hypothesis revealed that hysterics and dysthymics, and extraverts and introverts, did not differ in their scores on the test. The test itself consists of a box illuminated inside and having a variable speed photographic shutter fixed at one end. A picture is inserted in the box at the end opposite to the shutter and the subject is required to view through the shutter aperture for various exposure times and to report what he sees. A maximum of eleven exposures for each picture is given, these being arranged as follows: exposures 1 and 2 are given at a shutter speed of $\frac{1}{100}$ sec., 3 and 4 at $\frac{1}{5}$ sec., 5 and 6 at $\frac{1}{2}$ sec., 7 and 8 at 1 sec., 9 and 10 at 3 secs., and 11 at 30 secs. The subject's score on each picture is the number of the exposure at which he correctly recognizes the picture, unless he fails completely when he is given a score of 12. The total test score is the sum of the individual scores from each of the six pictures.

The mean score for each group on this test is shown in Table 2.13. Analysis of variance indicated that the overall differences between the groups were not significant, although it can be seen from the table that the group means are in the expected direction, the three patient groups showing generally longer recognition times than normal.

TABLE 2.13

Picture Recognition Test

	Mean	S.D.
Normals	39·31	7·732
Dysthymics	40·38	3·931
Hysterics	44·75	7·084
Schizophrenics	40·81	9·726

Body Sway Test. This is a test of primary suggestibility, shown by Eysenck (1947, 1953b), and on a number of other occasions subsequently, to be closely associated with neuroticism but not with extraversion. The particular apparatus used here consisted of a simple pulley arrangement which allowed a measure of forward and backward sway to be taken on a vertical scale graduated in inches. The suggestion that he was falling forward was made to the subject by means of a tape-recording of the experimenter's voice, the suggestion being continued for two minutes. Both total backward and forward sway were noted, subjects falling completely in either direction being given a sway measure for that direction of six inches. Total sway regardless of direction was also calculated for each subject. Mean sway for each of the four groups is shown in Table 2.14. As expected, the two neurotic groups were alike in showing the greatest amount of body sway, with normals showing least sway. Analysis of variance revealed, however, that these differences were not significant.

TABLE 2.14

Body Sway: total sway

	Mean	S.D.
Normals	2·05	1·319
Dysthymics	2·77	1·833
Hysterics	3·08	2·169
Schizophrenics	2·34	2·056

Sorting Test. Here the subject is asked to sort a number of objects in as many ways as possible and to give adequate explanations for each sorting method. Tests along these lines have been used for a number of years in the study of 'abstract' thinking (Vigotsky, 1934; Goldstein and Scheerer, 1941) and the general finding of impairment of concept formation in schizophrenics is too well known to require elaboration here. A thorough examination of the problem has been made by Payne (1954), whose sorting test was used in the present study and who gives a detailed description of its use in the investigation of concept formation. Briefly, the material used consisted of twelve objects of common geometrical shape, viz. four square, four circular, four triangular. All the objects were different, but could be grouped together in a variety of ways according to a number of characteristics. These included, in addition to shape, such qualities as size, thickness, material, weight, and colour (both hue and brightness). After a preliminary demonstration if the subject did not grasp the instructions, he was asked to sort the objects in as many ways as possible, giving a reason each time for the sorting method used. Two scores were taken on the test, viz. the number of correct sortings and the total time taken, the latter being converted to a log. time value.

The group mean values for these two scores are given, respectively, in Tables 2.15 and 2.16. It can be seen that schizophrenics are, as predicted, poorest on both measures derived from this test, showing fewer correct sortings and a longer time than the other three groups. The differences between the groups were not, however, statistically significant on analysis of variance.

TABLE 2.15

Sorting Test: number of correct sortings

	Mean	S.D.
Normals	4·25	1·479
Dysthymics	3·56	1·972
Hysterics	3·50	3·969
Schizophrenics	2·56	2·123

TABLE 2.16

Sorting Test: log. time

	Mean	S.D.
Normals	2·5094	0·3059
Dysthymics	2·5032	0·2557
Hysterics	2·5745	0·2464
Schizophrenics	2·6701	0·2429

DISCUSSION OF GROUP DIFFERENCES

Neurotic and normal groups. From the overall group differences a number of consistent findings emerge which essentially confirm most of the hypotheses proposed. It is clear that, with the exception of a few measures, hysterics will differ from dysthymics in a systematic and predictable fashion and that, again with a few exceptions, normals will tend to fall midway between these two groups. Hysterics are found to have poor vigilance, a relatively short spiral after-effect, large negative time errors in the judgement of sound intensity, and a tendency to underestimate time intervals to a greater extent than dysthymics. Further, in so far as the hysteric group emerged significantly more extraverted than the dysthymic group the theoretical position adopted at the beginning of this paper seems to be substantiated. Some discrepancy arises, however, when comparisons are made between the normal group and the two neurotic groups, particularly between hysterics and normals. The attentive reader will recall the finding, reported earlier, that the hysteric group did not emerge more, but slightly less, extraverted than the normal group. This, of course, was not predicted. Taking this into account, it is apparent that the hysterics' performance is somewhat exaggerated, out of proportion as it were, to the degree of extraversion present in this group. A somewhat similar effect, but in reverse, was seen in the results of the vigilance test, where dysthymics and normals of equivalent degrees of introversion differed markedly in their fall-off on this test. Since the inhibition hypothesis proposed as a general explanation for the results seems to be fairly substantially confirmed, it is clearly necessary to account for these disproportionate inhibitory effects.

One possibility is that motivational differences of some kind were operating to produce these effects. A source of such differences between the hysteric and dysthymic groups, at any rate, might conceivably lie in the N-scale of the MPI, assuming neuroticism as measured by this scale to act as a drive. Sigal, Star, and Franks (1958a), for example, have reported that hysterics tend to have lower N-scale scores than dysthymics. In the present sample the difference between the two neurotic groups in N-score was certainly in this direction but was not significant. Inspection, however, of responses to individual items of the N-scale did reveal some differences between the two groups. It was found, for example, that to the question 'Would you rate yourself as a tense or high-strung individual?' out of sixteen subjects in each group twelve dysthymics, but only four hysterics, answered in the affirmative. Chi-square calculated between these frequencies had a value of 8·56 (d.f. 2), which is easily significant at the 0·02 level of confidence. While this particular finding is in itself very slender evidence it is perhaps not coincidental that it tends to confirm the clinical differences normally held to distinguish these two groups of neurotics. Traditionally, the true hysteric has been recognized as lacking in drive, or more specifically lacking in overt anxiety or tension; the 'belle indifference' accompanying this is well known. By contrast, the dysthymic is characterized by high levels of 'free' anxiety and tension. Although no evidence is forthcoming in the present study of the likely source of such differences, if they exist, in the sample used here, the work of van der Merwe (1948) has demonstrated some of the autonomic concomitants of what appears to be the same underlying variable, hysterics and dysthymics being shown by him to fall on either side of a normal group on a measure of emotional tension. It seems not unlikely, therefore, that differences in what we may justifiably subsume under the heading of 'drive' may have accounted partly for the inter-group differences. Further support for this will be presented later, but meanwhile attention must be turned to the probable effects of drive on the experimental findings.

It is quite apparent that the effect of drive is likely to be particularly important on tests such as vigilance and the pursuit rotor. The asymptote of performance reached on the latter, for example, is clearly determined to a large extent by drive and it is noticeable in this respect that hysterics performed at a markedly lower level than other groups on the pursuit rotor. Similarly, the low reminiscence scores of hysterics may be accounted for if we assume a lowered degree of drive in this group, since, as Eysenck (1956b) has pointed out, reminiscence is closely related to drive, because the lowered 'critical level' associated with it limits the individual's tolerance of inhibition and thus allows less to build up. One other consequence of changes in the critical level is that rest pauses during performance are likely to occur earlier and more frequently. The reader may be briefly reminded here of the process involved. The onset of reactive inhibition is expected to culminate in a cessation of behaviour when I_R equals the level of positive drive, that is at the critical level. Through the reduction of negative drive (I_R) the rest pause becomes a conditioned habit of not responding ($_SI_R$). The possible analysis of vigilance in these terms has already been discussed in considering this test and Broadbent (1958) has recently indicated the importance of rest pauses in determining vigilance performance. The poor vigilance performance of hysterics and the superior performance of dysthymics relative to normals of equivalent introversion would certainly be consistent with such an analysis.

An alternative possibility is that increases and decreases in drive result not simply in differential tolerance levels of inhibition, but serve to alter the individual's susceptibility to it. Such an analysis would accord well with a central view of drive, such as that of Hebb (1949), particularly the later modifications to his theory (1955), which identifies drive as a state of cortical arousal arising from the activity of the brain stem reticular formation. This interpretation of drive may be meaningful in the present context in so far as it would take account of a possible interaction between excitatory and inhibitory processes, both occurring at a central level. In terms of the present theory it could be thought of tentatively as a shift, possibly temporary, in the balance of excitation-inhibition proposed to underly the personality dimension of introversion-extraversion. A heightened state of drive or arousal in the dysthymic, for example, could be thought of as effectively a shift towards greater excitation or, conversely, lower inhibition. A corresponding lowering of the arousal level in hysterics could similarly account for the apparently disproportionate inhibitory effects seen in this group. Particularly relevant to this formulation of drive is the suggestion, considered above, that the drive differences under discussion may have an autonomic basis, since the level of arousal is almost certainly determined (Hebb, 1955) by the amount of internal as well as external stimulation.

Clearly, such an interpretation of drive does not essentially contradict the analysis of pursuit rotor and vigilance performance discussed above. On tasks of this kind drive in the Hullian sense may still be expected to have a limiting effect, via the critical level, on the level of inhibition reached before rest pauses begin. The opportunity, during such enforced rest pauses, partially to dissipate previously acquired inhibition may be expected to have a 'levelling' effect on measures such as reminiscence, in the manner discussed earlier. A further factor complicating the reminiscence measure is the role played by habit. As we shall see when the pursuit rotor performance of psychotics is discussed, the reminiscence measure is, in fact, a multiplicative function of both habit strength ($_{S}H_{R}$) and inhibitory potential (I_R). Any reduction in habit level, as the pursuit rotor curves of hysterics suggest may be the case in this group, will necessarily lower the actual reminiscence changes found, even though the inhibitory component has a relatively high value. (The reader is referred, for a fuller discussion of this point, to Treadwell, elsewhere in this book.) Indeed, it is possible that reminiscence scores will prove unsuitable as direct measures of inhibition, at least where variations in drive and/or habit are expected.

Despite the lack of certainty in the results and the consequent need for further investigation, it seems not unreasonable on the evidence available to conclude that differences in drive or 'arousal' level in neurotics may serve to exaggerate the effects of inhibition on performance which are predictable in normal subjects from extraversion alone. The possibility that an actual shift occurs in the excitation-inhibition balance, while according well with modern views of drive as a central process, can only be regarded as tentative. Nevertheless, it would seem important theoretically to determine exactly the extent to which the predicted relationship between extraversion and inhibition is altered by such factors as drive once we go beyond the middle ranges of normality. In order to clarify some of the issues emerging from the present findings further work is necessary, particularly in establishing more exactly the manner in which possible drive differences in hysterics and dysthymics may be paralleled by changes on autonomic and other measures. Supplementing this, direct comparisons, on tests of the kind used here, between normals and neurotics of equivalent levels of extraversion and introversion would also be expected to throw further light on the problem. Unfortunately, in the present study the numbers were too small to make such comparisons with any degree of confidence and this must be left to future research.

Psychotic group. At the beginning of this paper reference was made to two groups of variables which have been commonly thought in the past to differentiate the schizophrenic from normal. There is the possibility, it will be recalled, that the schizophrenic is characterized by excessive amounts of inhibition, or alternatively, that a lack of drive, or ability to sustain attention, is largely responsible for his poor performance on most tasks. Our own hypotheses directed at this group were, it must be admitted, formulated on the basis of an intuitive compromise between the two approaches. It will be apparent to the reader by now that the results for the psychotic group do not emerge unequivocally in support of either. In so far as the schizophrenics' performance on some tests is similar to that of hysterics, whom we have already held to be characterized by higher levels of inhibition than dysthymics, it can certainly be said that such results support the suggestion that high levels of inhibition also characterize the schizophrenic. The time error, time judgement, and vigilance tests, for example, could readily be analysed in these terms, although the results of the latter are also consistent with an explanation in terms of decreased motivation, or, again, as an interaction between drive and inhibition. Most critical, however, to an explanation in terms of excess inhibition alone are the low pursuit rotor reminiscence scores of schizophrenics and the finding that their spiral after-effect is, if anything, longer than normal. It is certainly possible that the reminiscence finding could be explained if we assume a lowered critical level due to lowered drive in this group. Such an explanation would then neither contradict nor confirm an analysis in terms of increased inhibition. However, the spiral results still present a difficulty.

Brief mention has already been made, however, of a suggestion by Venables (1959) that it is the rate of dissipation, rather than the absolute level, of inhibition which is important in the performance of schizophrenics. Such a concept has a number of advantages not shared by the two alternatives considered above and appears to merit further consideration. It is clear, for instance, that if it is in fact the rate of dissipation of inhibition which is critical, then the spiral result is no longer unusual, since the length of after-effect is uninfluenced by the speed at which the inhibition accounting for the phenomenon decays. It is in fact only the *degree* of inhibition induced which is critical, this inhibition dissipating in its own time, as it were, after the test has finished. The concept of differential rates of inhibitory decay has recently been applied to the pursuit rotor by Broadhurst and Broadhurst (unpublished paper), who report similar findings with regard to reminiscence in schizophrenics as those found here. Their suggestion is that the reason schizophrenics show little reminiscence is because the ten minutes rest allowed between phases is insufficient to allow inhibition in psychotics to dissipate. There are, however, a number of difficulties with this explanation, as the Broadhursts themselves point out. The main difficulty concerns the effects of conditioned inhibition. Since the extinction of conditioned inhibition ($_{S}I_{R}$) is dependent essentially on the fact that reactive inhibition (I_R) has dissipated during the ten-minute rest and can no longer reinforce the habit of not responding ($_{S}I_{R}$), then schizophrenics, not having dissipated their inhibition, should not show the effects in the post-rest learning curve suggestive of the extinction of this habit. In other words, the abrupt rise in the post-rest learning curve, said by Eysenck (1956c) to be the result of the extinction of conditioned inhibition, should not occur in schizophrenics. However, as can be seen from Fig. 2.6, there are in fact signs of an exaggerated, rather than a diminished, extinction increment in the learning curve (Phase II) of the psychotics used here, although the Broadhursts' results are somewhat dissimilar.

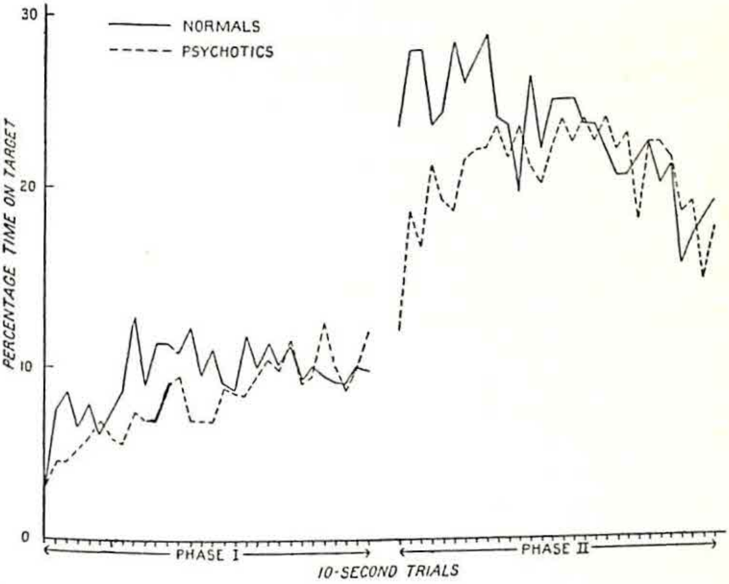


FIG. 2.6. Pursuit Rotor—Detailed Performance Curves for Normals and Schizophrenics in Phase I and II. Notable Features, discussed in the text, are the 'Straight Line' Learning Curve and Exaggerated 'Warm-Up' shown by Schizophrenics.

A reconciliation of the two apparently contradictory findings of the high $_{S}I_{R}$ and low reminiscence is possible, however, if we assume that the time intervals for inhibitory decay are shorter than that implied by the Broadhursts—i.e. greater than ten minutes—but still longer than that required by normals. It is clear, for instance, that if schizophrenics dissipate inhibition at a slower rate than normals then one consequence of this will be that the involuntary rest pauses occurring when I_R reaches the critical level will be somewhat prolonged, since a greater time will elapse between the onset of the rest pause and the resumption of practice following the decay of inhibition. Theoretically there would be two consequences of this. First, the amount of conditioned inhibition subtracting from the habit factor will be proportionally greater at any one point in learning than in normals. Secondly, the breaks in performance, if they are long enough, will reduce the amount of practice taken and thus retard the growth of habit strength. Both of these, and particularly the first, are likely to delay improvement in Phase I and may be critical especially at the beginning of this phase when ${}_{S}H_{R}$ values are low. This may incidentally account for the rather delayed learning curve shown by schizophrenics (see Fig. 2.6), although an equally meaningful explanation of this has been given by O'Connor and Rawnsley (1959), who suggest that the 'straight line' learning curves shown by some psychotics reflects the progressive arousal of drive in these subjects. Whatever the correct explanation of this aspect of the results, the position in any case would seem to be that at the end of Phase I schizophrenics are likely to have more ${}_{S}I_{R}$ than normal. Following the rest they would then be expected to show a large extinction increment.

Turning to the reminiscence effect the reader is referred to the following modified equation, suggested by Gwynne Jones (1958) as a more logical representation than that presented by Hull of the components determining effective reaction potential or performance level:

$${}_{S}\bar{E}_{R} = (D \times {}_{S}H_{R}) - (D \times {}_{S}I_{R}) - (I_{R} \times {}_{S}H_{R}) + (I_{R} \times {}_{S}I_{R})$$

On the final trial of Phase I performance is determined by all four components in this equation. On the first trial of Phase II it is determined by only the first two, the components falling out—because I_R has decayed in the normal way after ten minutes—being (*a*) the negative component ($I_R \times {}_{S}H_{R}$), this being the component that gives rise to the usual reminiscence effect; and (*b*) the positive component ($I_R \times {}_{S}I_{R}$), this *lowering* the reminiscence effect according to its magnitude. It is clear that the removal of this second component will, together with the subtracting influence of the still remaining negative component ($D \times {}_{S}I_{R}$), have proportionally greater effects in schizophrenics than in normals, if, as we have suggested ${}_{S}I_{R}$ has a higher value in the psychotics. The relatively good performance of the psychotic group would not in fact be revealed until this considerable amount of conditioned inhibition had been extinguished, and the reminiscence increment would be abnormally low. One other point should be made. From Fig. 2.6 it can be seen that in fact *no* reminiscence occurs if we take the last and first trials of Phases I and II as our reference points, which, it will be remembered, was not done in taking the reminiscence measures here. (In fact, of course, the performance measure taken on trial 1 of the second phase is also affected to some extent by the necessity for the subject to warm up, suggested by Eysenck (1956c) to last about 10–20 seconds. This will subtract from the 'true' reminiscence effect although in normal subjects it can be regarded as of little significance when compared with the magnitude of the reminiscence changes usually found. In psychotics, however, if our previous analysis is correct, the subtraction of such a component is likely to result in a zero or even negative change in performance following a rest, particularly if, as seems likely, schizophrenics are slower than normal in warming up.)

While the interpretation given here is necessarily *post hoc* it does have the merit of suggesting further testable hypotheses and does, furthermore, account for all the results more consistently than the other possibilities considered. In vigilance, for example, a lower detection rate of signals would be the expected result of slow inhibitory decay. Indeed, any prolongation of the time spent away from tests of this type may be regarded as particularly critical, since the probability of signal detection during these periods is reduced to zero. A higher growth rate or absolute level of inhibition may, in fact, be less critical since the inhibition accumulating during vigilance tasks arises largely as the result of stimuli (in our case digits) which are irrelevant and contribute in no way to the subject's score. This may account for the slight inferiority in vigilance of schizophrenics compared with hysterics, in whom inhibitory growth rate rather than decay rate has been more frequently emphasized. It is certainly true that to a greater extent on some tests than others the three aspects of the inhibitory process—absolute level, growth rate, and speed of decay—may appear to have effectively identical or similar effects on measures of performance. This is exemplified by the greater tendency generally for the performance of schizophrenics to resemble that of hysterics rather than dysthymics, the notable exception to this being the spiral after-effect measure. By means of suitably constructed tests and experimental designs aimed at testing specific hypotheses it should, however, be possible in future research to isolate the relevant aspects of inhibition which differentiate the schizophrenic from the hysteric and from other normal and abnormal personalities.

THE OVERALL ANALYSIS OF RESULTS

So far we have been concerned with the analysis and discussion of individual tests in terms of group differences in behaviour. In this form the data provided, with a number of provisos, fairly substantial confirmation of the theoretical framework within which the original hypotheses were formulated. Now, however, we must turn to a consideration of the relationships between tests and attempt an analysis of

TABLE 2.17

Key to measures included in First and Second Analyses

Number	Test	Measure and description
0	SSG (Intelligence) rating	Sign reflected in final factor loadings.
1	Vigilance	First part: total score.
2	Vigilance	First part: regression coefficient '*b*' or 'starting level'.
3	Vigilance	First part: regression coefficient '*a*' or 'improvement'. Sign reflected in final loadings.
4	Vigilance	Second part: total score.
5	Vigilance	Performance change from first to second part of test.
6	Time Error	Mean of error at four S–V intervals.
7	Sorting Test	Number of correct sortings.
8	Sorting Test	Log. Time.
9	Picture Recognition	
10	Body Sway Test	Total sway in both directions.
11	Time Judgement	Mean error of underestimation for four time intervals.
12	Spiral	Mean of four trials.
13	AFF	
14	Pursuit rotor	RS1.
15	Pursuit rotor	RS2.
16	Pursuit rotor	% RS1 or reminiscence change as a percentage of pre-rest performance level.
17	Pursuit rotor	% RS2.
18	Pursuit rotor	Phase I: regression coefficient '*b*' or 'performance level'.
19	Pursuit rotor	Phase I: regression coefficient '*a*' or 'rate of improvement'.
20	Pursuit rotor	Phase II: regression coefficient '*b*'.
21	Pursuit rotor	Phase II: regression coefficient '*a*'.
22	Pursuit rotor	Phase III: regression coefficient '*b*'.
23	Pursuit rotor	Phase III: regression coefficient '*a*'.
24	MMPI	HY Scale score.
25	MMPI	PT Scale score.
26	MMPI	PD Scale score.
27	MMPI	SC Scale score.
28	MPI	E-scale score.
29	MPI	N-scale score.

the main factors underlying them. Since the principal theme running through this investigation has been that individual differences on the wide variety of measures employed are explicable in terms of variations in underlying inhibition, then it would be expected that an 'inhibitory factor' would emerge from the intercorrelations between the tests and, further, that this would be identifiable with the behavioural dimension of extraversion. Of particular interest in this respect was the possibility of finding a link between inhibition and satiation according to the differential method of proof discussed earlier.

TABLE 2.18

Unrotated factor loadings for First Analysis (48 *subjects*)

Test	Factor I	II	III	IV
0	−0·49	−0·35	−0·41	−0·06
1	0·34	0·48	0·21	0·64
2	0·29	0·38	0·74	0·06
3	−0·05	−0·03	−0·72	0·50
4	0·63	0·14	0·19	0·18
5	0·24	−0·31	0·41	−0·56
6	0·13	0·36	−0·02	−0·02
7	0·28	0·26	0·16	−0·13
8	−0·08	−0·01	0·14	−0·01
9	−0·21	−0·06	−0·02	−0·25
10	−0·36	−0·05	0·04	−0·07
11	0·45	0·22	0·28	−0·14
12	−0·07	0·54	−0·44	0·09
13	0·46	0·04	0·11	−0·33
14	0·43	0·01	−0·38	−0·53
15	0·42	−0·18	−0·25	−0·22
16	0·07	−0·26	−0·26	−0·32
17	−0·18	0·04	0·41	−0·08
18	0·44	0·43	−0·10	0·32
19	0·12	−0·01	−0·04	−0·21
20	0·73	0·43	−0·23	0·03
21	−0·56	−0·32	0·03	0·09
22	0·73	0·16	−0·47	−0·04
23	−0·55	0·03	0·42	0·34
24	−0·31	0·67	0·06	−0·18
25	−0·46	0·74	−0·08	−0·28
26	−0·23	0·62	0·07	−0·37
27	−0·48	0·65	−0·06	−0·38
28	0·38	−0·56	0·07	−0·22
29	−0·33	0·49	−0·38	−0·16

In order to investigate this, two principal component analyses[1] were carried out on the data, correlations being calculated between thirty selected measures. These measures are indicated in Table 2.17, numbered to correspond with the entries in the tables of factor loadings, and briefly described where appropriate. The first analysis was carried out on the scores for the hysteric, dysthymic, and normal groups only, there being 48 subjects in this case. The same measures from all 64 subjects were then included in a second analysis. Four components were extracted from the correlation matrix of each analysis. The elements of the latent vectors appropriate to each component are considered here in their factor-loading form, the loadings for the first and second

---

[1] The writer is indebted to Mr. Benedict Nixon of the Psychology Department who carried out these analyses for him on the Ferranti Pegasus Electronic Computer.

TABLE 2.19

Rotated factor loadings for First Analysis (48 *subjects*)

Test	Factor				
	I	II	III*	IV*	h^2
0†	0·73	0·0	0·0	−0·07	0·538
1	0·57	−0·06	0·13	−0·67	0·794
2	0·79	0·12	−0·36	−0·12	0·782
3†	0·45	0·20	−0·56	0·46	0·768
4	0·60	−0·28	0·18	−0·07	0·476
5	0·25	−0·13	−0·32	0·67	0·631
6	0·25	0·22	0·19	−0·07	0·152
7	0·40	−0·12	0·08	0·09	0·189
8	0·02	0·04	−0·15	0·0	0·024
9	0·19	0·16	−0·10	0·20	0·112
10	−0·25	0·17	−0·22	0·0	0·140
11	0·57	0·0	0·05	0·15	0·350
12	−0·05	0·41	0·50	−0·28	0·500
13	0·40	−0·07	0·14	0·39	0·337
14	0·09	0·0	0·52	0·58	0·615
15	0·07	−0·26	0·32	0·36	0·304
16	−0·21	−0·11	0·17	0·40	0·245
17	0·12	0·15	−0·41	0·02	0·205
18	0·44	−0·02	0·42	−0·34	0·486
19	0·05	0·02	0·08	0·22	0·058
20	0·58	0·05	0·67	−0·01	0·788
21	−0·51	0·0	−0·38	−0·10	0·414
22	0·32	−0·22	0·78	0·14	0·779
23	−0·13	0·15	−0·59	−0·44	0·581
24	0·12	0·74	0·0	−0·13	0·579
25	−0·03	0·91	0·08	−0·10	0·845
26	0·16	0·74	0·02	0·07	0·578
27	−0·07	0·89	0·02	0·02	0·798
28	0·05	−0·53	−0·05	0·48	0·516
29	−0·22	0·60	0·31	−0·09	0·513

* Indicates that these factors have been reflected (see text).
† Indicates that these tests have been reflected (see text).

analyses being shown in Tables 2.18 and 2.20 respectively. In order to achieve psychological meaning from the components orthogonal rotations were performed on both sets of data. The rotated factor loadings for the analysis carried out on 48 subjects are shown in Table 2.19 and the corresponding values for the second analysis in Table 2.21. It should be noted that in Table 2.19 the signs of all loadings for Factors III and IV have been reversed to facilitate interpretation and the comparison between the two analyses. Similarly, in both Tables 2.19 and 2.21 the signs for tests 0 and 3 have been reflected, again to facilitate interpretation. Test 0 (the SSG or intelligence rating) was, it may be remembered, scored in such a way that a high SSG referred to low intelligence level. In the case of Test 3 the regression coefficient 'a' is more appropriately considered in the context of the vigilance test as a measure of rate of fall-off, rather than of rate of improvement. The analysis carried out on the normal and neurotic groups only will be considered first and the reader is referred again to Table 2.19.

The first factor extracted can fairly clearly be identified as a factor of intelligence, the loading of the intelligence measure itself (Test 0) being 0·73, and the Sorting test (Test 7) having, as might be expected, a loading of 0·40. Although this factor is not in itself of immediate interest it should be noted that a number of other measures, not so obviously associated with intelligence, also have relatively high loadings on this factor. Particularly noticeable in this respect are some of the vigilance measures (Tests 1–5 inclusive). Vigilance starting-level (Test 2), for instance, has a higher loading on this factor than any other test, viz. 0·79. One possible explanation of this is that within the relatively wide range of intelligence covered here the inability of the duller subjects to grasp immediately the concept of odd versus even numbers was picked out in their poor performance at the beginning of the test. It is possible, therefore, that intelligence influences vigilance performance only up to a certain level of brightness but not beyond it. The Time Judgement test (Test 11) can also be seen to have a moderately high loading of 0·57 on the intelligence factor and zero or negligible loadings on the other three factors. Equally disappointing is the Time Error test (Test 6), the highest loading of which is 0·25, again on intelligence. Auditory Flutter Fusion (Test 13) seems also to be partly determined, in this sample at least, by intelligence, its highest loading (0·40) being on this factor. It may be significant that the last three tests have as an element in common the fact that they demand from the individual a degree of judgement, which, intuitively at least, one might expect to be subject to intellectual influences. Speaking generally, it may be true that some of the high loadings on Factor I of tests not obviously related to intelligence are due in part to the range of ability present in the sample, a number of subjects being somewhat duller than those normally employed in psychological experiments.

Factor II can also be interpreted fairly clearly. It is mainly determined by the questionnaire measures from the MMPI and MPI (Tests 24–9 inclusive) and is similar to that reported, for example, by Eysenck, Granger, and Brengelmann (1957), who called their factor 'introspective abnormality'. In the present case it may be more specifically identified as 'neuroticism', most of the tests with high loadings on the factor being measures of this dimension. Few of the objective tests have appreciable loadings on the factor, although the Spiral test (Test 12) with a loading of 0·41, is a noticeable exception. This test can also be seen to have loadings on the third and fourth factors and would seem, therefore, in view of our predictions about the test, to merit further discussion. This will be delayed, however, until the other factors have been described.

Factor IV is most conveniently considered next as this too seems to be fairly well defined. In fact, it may be interpreted with some confidence as the factor of 'extraversion' which we predicted would emerge from the analysis, the loading of the E-scale (Test 28) being 0·48. Its identification with an underlying factor of 'inhibition' is supported by a number of loadings on the objective tests. The loadings of −0·67 and 0·46 shown by Tests 1 and 3, for example, indicate that poor overall vigilance performance and rapid fall-off on this test, both indicative of high levels of inhibition, are positively associated with extraversion. Similarly Test 5 (Vigilance change) has a loading of 0·67 on this factor and confirms our hypothesis that disinhibitory effects will be greater among extraverts, although the moderate loading of this measure on the third factor should also be noted. Turning to the other tests RS1 (Test 14) and RS2 (Test 15) have loadings of, respectively, 0·58 and 0·36, reminiscence, as predicted, being higher in extraverts, although again there are appreciable loadings on Factor III. AFF (Test 13) has a loading of 0·39 on the factor, but the Time Error and Time Judgement tests, with loadings of −0·07 and 0·15 respectively, fail to confirm our expectations that the group differences found on these two tests would be explicable in terms of the link between inhibition and extraversion. Indeed, the very low communality of Time Error (Test 6) indicates the very small extent to which the four factors extracted account for performance on this test. Partial confirmation of the similarity between satiation and inhibition is forthcoming in the loading on extraversion of the Spiral test, but the value of −0·28 is noticeably lower than the same test's loading on other factors, particularly Factor III to which we now turn.

Interpretation of the third factor is less certain than in the case of the other three factors. It has loadings of −0·56 on Vigilance Fall-off, 0·52 and 0·32 on RS1 and RS2 respectively, 0·50 on the Spiral, and 0·42, 0·67, and 0·78 on the performance level measures for the three phases of the pursuit rotor experiment. In addition, the N-scale of the MPI

(Test 29) has a moderate loading of 0·31. The pattern of loadings suggests that the most likely interpretation is that this is a 'drive' or 'motivational' factor. Undoubtedly supporting this are the high positive loadings of pursuit rotor performance level, of which one of the major determinants is certainly drive. Similarly, low degrees of fall-off in vigilance may be meaningfully interpreted as being closely associated with high drive levels, while, as we have already seen, the latter is also likely, other things being equal, to result in increased amounts of reminiscence. Finally, the moderate loading of the N-scale on this factor lends support to the interpretation, in so far as neuroticism as measured by this scale can be said to have drive properties. It is interesting in this respect that the MMPI measures determining Factor II, which we have interpreted as a factor of general neuroticism, have virtually zero loadings on Factor III. It is just possible that the N-scale taps something more specific, such as anxiety. This would certainly link up the generally lowered N-scale scores and poorer objective test performance of hysterics with the suggestion, made earlier, that symptomatic anxiety may be one of the important determinants of the dysthymic-hysteric differences found here.

The probability that Factor III can be meaningfully interpreted as one of drive certainly lends general support to an hypothesis that variations in motivation may have accounted for some of the intergroup differences reported and it is interesting that a factor of this kind should have emerged which confirms the previous interpretation of the individual test results.

One relevant aspect of the analysis requires further comment and that is the result of the Spiral test. It can be seen that Factors III and II provide the highest loadings for this test, its loading on extraversion being somewhat lower than our original hypotheses would have led us to expect. In other words, in this sample increases and decreases in drive may be said to have had the effect of, respectively, lengthening and shortening the spiral after-effect over and above the latter's negative correlation with extraversion. One possible explanation of this is that increased motivation, in terms of greater task-involvement, induces the subject to fixate more tenaciously, while, conversely, low motivation causes him to indulge in eye and head movements,[1] which, because they will produce frequent shifting of the fixation point, are likely to result (Holland 1957) in a reduction of the after-effect. In other words, since it is known (Holland 1958) that, within wide limits, the length of spiral after-effect increases directly with the amount of previous stimulation, then it is possible that reduced drive lowers the after-effect simply by lowering the degree of stimulation giving rise to it. On the other hand, it is likely that particularly gross and frequent movements, which remain to be demonstrated, would be necessary to account for the magnitude of the effect.

---

[1] An interesting speculation, worthy of further investigation, is that these eye movements are the parallel at the perceptual level of the spontaneous rest pauses occurring in continuous work. If so, then an exact determination of the role played by drive in perceptual tasks would be of particular importance.

Another possibility is that the loading on Factor III simply confirms the suggestion made earlier that drive, in so far as it may be identified as a central excitatory process, would be expected to alter in some unspecified way the balance between excitation and inhibition which we have assumed to underlie the spiral after-effect phenomenon. It is, however, impossible on the present evidence to decide which of the various possible explanations is correct.

TABLE 2.20

Unrotated factor loadings for Second Analysis (64 *subjects*)

Test	Factor			
	I	II	III	IV
0	–0·66	–0·25	0·33	–0·03
1	0·71	0·29	–0·15	–0·39
2	0·70	0·26	–0·35	0·02
3	–0·18	–0·03	0·37	–0·57
4	0·75	0·10	–0·27	–0·06
5	0·12	–0·20	–0·49	0·64
6	0·42	0·15	0·07	–0·21
7	0·54	0·25	–0·18	0·13
8	–0·13	0·04	–0·41	0·13
9	–0·30	0·05	–0·02	0·11
10	–0·19	0·13	0·12	–0·12
11	0·35	–0·08	0·04	0·0
12	–0·03	0·30	0·35	–0·37
13	0·48	–0·07	0·06	0·22
14	0·24	–0·19	0·58	0·44
15	0·37	–0·29	0·45	0·16
16	–0·17	–0·27	0·36	0·34
17	–0·15	0·01	0·03	0·26
18	0·53	0·25	0·0	–0·07
19	0·18	0·02	–0·37	0·03
20	0·72	0·17	0·12	0·09
21	–0·60	–0·20	–0·36	–0·20
22	0·70	–0·14	0·32	0·06
23	–0·56	0·18	–0·56	–0·18
24	–0·05	0·74	0·10	0·13
25	–0·28	0·86	0·10	0·21
26	–0·20	0·62	0·05	0·44
27	–0·42	0·75	0·11	0·30
28	–0·11	–0·57	–0·13	0·39
29	–0·19	0·57	0·24	0·03

The second analysis carried out on all 64 subjects reveals, as can be seen from Tables 2.20 and 2.21, a very similar factor pattern, although there are a number of discrepancies worth noting. All four factors, with the possible exception of Factor III emerge equally clear-cut as in the first analysis. As before, the intelligence rating (Test 0) has a high loading of 0·77 on Factor I, again supporting the interpretation that this is an 'intelligence' factor. As expected, the introduction of the psychotic group has the effect of increasing the loading of the Sorting test (Test 7) on the first factor from 0·40 to 0·61. The loadings of this test on the other factors are, however, virtually unchanged. As before, the vigilance

TABLE 2.21

Rotated factor loadings for Second Analysis (64 *subjects*)

Test	Factor I	II	III	IV	h^2
0*	0·77	0·04	0·0	0·0	0·594
1	0·75	−0·14	0·01	−0·42	0·759
2	0·83	0·04	0·0	0·0	0·690
3*	0·32	0·26	−0·02	0·56	0·484
4	0·78	−0·15	0·07	−0·01	0·636
5	0·25	0·12	−0·10	0·78	0·695
6	0·38	−0·10	0·15	−0·28	0·255
7	0·61	0·13	0·11	0·10	0·411
8	0·07	0·13	−0·35	0·24	0·202
9	−0·24	0·18	0·13	−0·03	0·108
10	−0·18	0·10	−0·04	−0·19	0·080
11	0·26	−0·16	0·19	0·01	0·129
12	−0·09	0·07	0·11	−0·57	0·350
13	0·36	−0·07	0·34	0·17	0·279
14	−0·09	0·0	0·76	0·21	0·630
15	0·05	−0·26	0·61	0·08	0·449
16	−0·38	0·0	0·36	0·26	0·342
17	−0·15	0·25	0·03	0·16	0·111
18	0·53	0·03	0·18	−0·20	0·354
19	0·32	−0·02	0·26	−0·06	0·174
20	0·62	0·0	0·43	−0·08	0·576
21	−0·42	−0·10	−0·46	0·40	0·558
22	0·42	−0·27	0·60	0·02	0·610
23	−0·20	0·20	−0·78	0·0	0·688
24	0·12	0·69	0·05	−0·30	0·583
25	−0·04	0·88	−0·03	−0·28	0·855
26	−0·02	0·78	0·07	0·03	0·615
27	−0·20	0·88	−0·04	0·16	0·842
28	−0·20	−0·25	0·03	0·64	0·513
29	−0·10	0·53	0·09	−0·34	0·415

* Indicates that these tests have been reflected (see text).

measures, particularly Tests 1, 2, and 4, are prominent in determining this factor, the loadings being somewhat higher than in the previous analysis. Also somewhat higher are Time Error (Test 6) with a loading here of 0·38 and the three measures of pursuit rotor performance level (Tests 18, 20, and 22) with loadings of 0·53, 0·62, and 0·42 respectively. These high loadings on Factor I are almost certainly due to the fact that reduced effective intelligence and poor performance on objective tests, such as pursuit rotor and vigilance, are more closely bound up with severity of illness in psychotics than is the case in other psychiatric groups.

Factor II can again be interpreted as a 'questionnaire neuroticism' factor, the loadings of the MMPI scales (Tests 24–7 inclusive) and the N-scale (Test 29) being almost identical with those in the first analysis. The E-scale (Test 28) loading, however, is reduced from —0·53 to —0·25. No objective test has a loading greater than 0·30 on this factor, the loading of the Spiral being reduced from 0·41 to virtually zero. It is interesting that, contrary to expectations, the Body Sway test (Test 10) had low loadings throughout both analyses, the loadings expected on Factor II being only 0·17 and 0·10 in the first and second analyses respectively. There seems no evidence, therefore, in the present investigation for Body sway as a test of general neuroticism. Picture Recognition comes off equally badly, the factors in both analyses accounting for only a small part of the communalities.

Factor III is less easily identified with the corresponding factor in the first analysis, its interpretation as a 'drive' component being somewhat more tenuous. The previous loadings on this factor of the vigilance measures, particularly Vigilance Fall-off (Test 3), are reduced virtually to zero, as is the loading of the N-scale. High loadings remaining, however, are those of the two reminiscence measures (Tests 14 and 15) with values of 0·76 and 0·61 and two of the pursuit rotor level measures (Tests 20 and 22) with loadings of 0·43 and 0·60. The loadings of AFF (Test 13) is increased from 0·14 to 0·34, and that of Sorting Log. Time (Test 8) from 0·08 to —0·35, although both of these are rather difficult to interpret. In so far as a number of measures have consistently high loadings in both analyses, there seems some justification for retaining the interpretation of this factor as one of 'drive'. It is not unexpected that the inclusion of the psychotic group, which accounts for one quarter of the sample, should have altered some of the relationships between tests. Apart from the sample as a whole being less homogeneous than previously, the effect is further exaggerated by the wide differences found on most measures within the psychotic group itself. Indeed, it is encouraging that the two analyses should have diverged so little.

Factor IV can, as before, be identified as one of 'extraversion', its loading on the E-scale (Test 28) being 0·64, which is somewhat higher than previously, although its loading on the N-scale is also increased from near-zero to −0·34. Of the objective tests the vigilance measures, Tests, 1, 3, and 5, retain their high loadings on this factor, while the Spiral (Test 12) emerges with a more clear-cut relation to extraversion. Its loading of −0·57 is over twice as high, while the loadings noted previously on the other factors are reduced considerably. The most marked discrepancy here is the reduced influence of reminiscence in determining the factor, the previously high loadings of RS1 and RS2 (Tests 14 and 15) decreasing to 0·21 and 0·08 respectively. This finding is perhaps not surprising when it is recalled that the overall level of reminiscence is reduced in schizophrenics, an effect which is not unlikely to mask the relationship of this measure with extraversion. It possibly confirms, as suggested earlier, that the similarity between the schizophrenic's performance and that of the extravert is to some extent a function of the measures we take, while fundamentally two different processes or two different aspects of the inhibitory process are operating.

To summarize the results of the two analyses, it can be said that, as predicted, some of the individual differences found in the group comparisons can fairly confidently be attributed to a personality factor identifiable as extraversion. Furthermore, it is true to say that this factor bears a systematic relationship to performance on a number of objective tests, in which central inhibition can reasonably be held to be a relevant underlying variable. It may be concluded from this, therefore, first, that inhibition as an explanatory concept has a general applicability to a number of different measures of behaviour; and secondly, that extraversion is an important behavioural concomitant of this underlying inhibition. A second important feature of the analysis is the evidence forthcoming that differences in what has loosely been called drive appear to play some part in determining performance on a number of tests. This, it has been argued, arises principally from differences on this variable between the two neurotic groups, who form one half of the sample. The demonstration of such a factor tends to confirm the possibility that some of the group differences in performance are accountable for in terms of an effective shift in neurotics in the balance of excitation-inhibition underlying introversion-extraversion.

Finally, turning to the rather less positive findings reported in this section, a number of other important facts emerged which tended to be masked by the group comparisons. For example, despite very marked group differences exactly in the direction predicted, Time Error shows little relationship with extraversion, or indeed with any of the other factors extracted. Either time error is not mediated by the inhibitory process in the way we suggested or our results failed to reveal the effect for some other reason. Before it would be possible to draw any definite conclusion it would be necessary to repeat the experiment. Such a repetition would seem worthwhile in view of the apparent relationship between time error and personality factors. A similar conclusion applies to Time Judgement, where the significant correlation found in the normal group between extraversion and time estimation suggests a real relationship in need of further investigation. To some extent the opposite trend was true of the Auditory Flutter Fusion. Here the difference between hysterics and dysthymics was negligible but some relationship with extraversion emerges from the overall analysis. It is, however, in the direction opposite from that predicted and from the results gathered on normals by Holland and reported by him elsewhere in this book. As we saw in discussing this test, it seems not to discriminate between the two types of neurotic but rather between neurotics as a group and normals, the latter achieving higher AFF scores. Since the neurotic groups taken together are less extraverted than the normal group, it is inevitable that positive, rather than negative, correlations should occur between AFF score and extraversion. Nevertheless, it is a discrepancy requiring an explanation which the present results fail to provide. Loadings on intelligence and, in one case, on drive are difficult to interpret in a meaningful fashion, although a negative correlation with motivational factors is not irreconcilable with the lowering of attention which we might expect to be present in some neurotics and psychotics and to be accompanied in these subjects by poor discrimination on tests of this kind. Whatever the correct interpretation, discrepancies such as these clearly suggest the need for further investigation into what appears to be an important area of personality research.

SUMMARY AND CONCLUSIONS

An attempt has been made to examine the validity of Eysenck's 'inhibition hypothesis' when applied to a variety of tests falling broadly into the categories of perception, attention, and learning. The major aim of the investigation was to test specific hypotheses arising from Eysenck's theory that groups of dysthymic, hysteric, and normal subjects would behave on these tests in a manner predictable from their respective positions on the personality dimension of introversion-extraversion. Comparisons between the behaviour of these groups and that of a group of schizophrenics also included in the design were expected to suggest the extent to which the behaviour of psychotics was explicable within the terms of Eysenck's theory or suitable extensions of it. In addition, an examination of the interrelationships between test results gathered on the same population would, it was anticipated, indicate the generality of the inhibition hypothesis as an explanatory concept for extraversion.

Results from the tests taken individually confirmed, with very few exceptions, the hypotheses proposed in each case, hysterics and dysthymics behaving in a manner suggestive of, respectively, high and low levels of central inhibition. Normal subjects tended to fall between these two groups, while the performance of psychotics was more like that of hysterics than that of any other group. Thus, in so far as the various hypotheses were confirmed, Eysenck's theoretical position could be said to have been vindicated. However, the relatively low, and unpredicted, level of extraversion in the hysteric group, and a number of aspects of the performance of dysthymics, indicated what appeared to be an exaggeration in the neurotic groups of the inhibitory effects predictable alone from the degree of extraversion evident in these subjects. The possibility was discussed that drive differences between hysterics and dysthymics may have accounted for the effect. The results for the psychotic group tended to support the view that a slow rate of inhibitory decay may be more characteristic of the schizophrenic than the high levels of (protective) inhibition more often invoked in studies of schizophrenia. Despite some support from other recent studies such a conclusion must be regarded more as a source of further testable hypotheses than as a statement of theory.

In order to investigate the interrelationships between the various tests, two principal component analyses were carried out on thirty measures derived from the objective tests and from personality questionnaires. The first analysis utilized the normal and neurotic groups only, while all subjects were incorporated in the second analysis. Both agreed in demonstrating four meaningful factors, the first two being identifiable in each case as 'intelligence' and 'questionnaire neuroticism'. Of the remaining two, one could be fairly clearly identified as a factor of 'extraversion'. This was found to be related, as predicted, to a number of measures from the objective tests, thus lending further support to the generality of Eysenck's hypothesis. Support was also found for the suggestion that motivational influences may have accounted for some of the inter-group differences, a factor being present which was identifiable as one of 'drive'.

A number of conclusions may readily be drawn from the results as a whole, but perhaps the most interesting, and certainly the most important, aspects of the investigation are the questions posed by it rather than the answers it provides. It is encouraging, but not unexpected in view of previous research, that many of our hypotheses should have been confirmed or that further evidence should have been found of a link between extraversion and central inhibition. Clearly, however, one of the major questions raised by the results concerns the possible influence on performance, and perhaps, therefore, on the inhibitory processes underlying performance, of factors other than extraversion. By selecting neurotic groups in an attempt to represent the extremes of extraversion-introversion it seems likely that unwittingly we introduced, as the evidence in fact suggests, other important variables. Of these the most relevant is certainly the drive variable isolated in the analysis of results. The implications inherent in this finding seem to have important bearing on the link that has been made between extraversion and inhibition, since there seems to be a real possibility of an actual or effective shift in inhibition and excitation levels in neurotics. Differences at the autonomic level between dysthymics and hysterics may be meaningful in this context, since a link is then feasible between cortical inhibition and drive as a central arousal or excitatory process.

Probably more questions than answers were forthcoming from the results for the psychotic group, although the general similarity between hysterics and schizophrenics on many tests seems to be one important finding which emerges. While not in itself unexpected, it does, nevertheless, help to delimit the area for future research. It is, of course, the differences rather than the similarities between hysterics and psychotics which are of greatest importance and the possibility of identical effects on performance of what may be different aspects of the inhibitory process suggests the need in future studies for carefully planned experimental designs and techniques.

In conclusion, it may be said with some confidence that the results of the investigation as a whole undoubtedly bear witness to the need for the experimental study of personality to be carried out within a theoretical framework which is not only capable of generating testable hypotheses, but which is also modifiable in the light of new facts emerging from the investigation of these hypotheses. That this has not always been the case in personality research in the past is an explanation, if not a justification, for some of the uncertainties apparent in the findings presented here.

THE EXCITATION-INHIBITION BALANCE IN NORMALS

(This study was made possible by a grant
from the Medical Research Council, to
whom we wish to express our thanks.)

3

MEASURES OF LEARNING AND CONDITIONING

R. A. Willett

GENERAL INTRODUCTION

THE investigation being reported in this chapter was initially planned as an attack on the problem of generality amongst various measures obtained from different conditioning procedures. An additional question of central moment was that of how regularly different conditioning measures related to the personality parameters introversion/extraversion and neuroticism. It will be recalled that a good deal of research has been carried out into the problem of how eyeblink conditioning relates to these two parameters by Eysenck and by his associates (Franks, 1956; Franks, 1957a; Franks, 1957b; Eysenck, 1957). It had been hoped to extend that research in this present investigation by utilizing conditioning techniques other than eyeblink conditioning. In fact, this aim has been only partly realized because on closer acquaintance with what were techniques new to the investigators, their status as conditioning processes seems somewhat dubious as will become clear in the later discussion of the results.

The basic conception related introversion/extraversion to a postulated neurological substruct, the particular balance of inhibition/excitation tendencies characterizing the central nervous system of a particular individual. It was suggested that a state in which inhibitory tendencies are apt to predominate characterizes the extravert while the obverse characterizes the introvert. Now clearly the most parsimonious corollary to this hypothesis would allow of only one mode of activity for such a substruct. In other words the simplest view of the hypothesis would be that characteristic properties of particular nervous systems would manifest themselves in the same way whenever the appropriate conditions presented themselves. Translating this argument, rather freely, into behavioural terms, we would expect a significant degree of correlation amongst different conditioning measures and further, we would expect all such measurements to be correlated significantly with a reliable measure of introversion/extraversion.

The terms of the initial research were broadened at a later date to incorporate processes other than conditioning. The notion of generality, implicit in the argument applied to conditioning, was extended to other test situations in which the inhibition/excitation balance seemed to play an important role (see chapter by Holland). The expectation then became that all relevant measures would correlate together and with extraversion. In view of the expansion of the original study, this wider aspect of generality, which has become associated with this investigation, has been given a separate treatment by Eysenck in Chapter 5. In that chapter the results of a principal component analysis, carried out on the whole matrix of intercorrelations of all tests administered in the work being reported here and in that reported by Holland, will be presented, together with some matters of the widest implications.

We shall here be concerned with the data from the 'conditioning' procedures without making specific reference to the problem of generality. We shall centre our interests on a number of crucial correlation co-efficients which have a direct bearing on expectations derived from the basic hypothesis about extraversion and the inhibition/excitation balance and we will not anticipate the 'second-order' type of analysis presented in the later section.

Because some rather new techniques were to be employed and, because some were new to us, a pilot study preceded the main investigation. Although we do not intend to give a full account of this we shall be referring to it from time to time and shall therefore include in our description of the experimental details any additional information necessary to outline this pilot study. In both the pilot and the main investigations, on the basis of the hypothesis connecting the inhibition/excitation balance with introversion/extraversion, it should be remembered, we expected that appropriate 'conditioning' measures would themselves intercorrelate and would correlate with a measure of this personality dimension. In point of fact our primary concern, in the pilot study, was to try out the unfamiliar technique and to examine only in a preliminary way, their relationship with introversion/extraversion.

SUBJECTS

Twenty male students from the University of London's Institute of Education formed the subjects of the pilot study. Of these, three had to be rejected through apparatus failure. The mean age of the remaining seventeen was 29·17 ± 6·28 years. These subjects answered an appeal for volunteer subjects made to Institute classes and received their expenses and a fee for their attendance.

The subjects of the main investigation were 100 students from the apprentice training school of a large industrial concern. Students at the school are selected by standardized test procedures and on the basis of their general school record. They are subdivided into two groups: a 'craft' group, whose members undergo a training in a skilled trade, and a 'student' group whose members study to qualify in some branch of mechanical or electrical technology. Both sub-groups are represented in our subject sample. Their ages ranged between sixteen and nineteen years. On the National Institute of Industrial Psychology's Test 90A they scored a mean of 82·2 ± 13·71 which, comparing these scores to those of an unselected group of secondary school children, places them towards the top of the middle 40% of pupils. In comparison to a Grammar school population they are placed somewhat below the middle of the same 40%. In both investigations, subjects with obviously relevant physical defects—e.g. any degree of deafness, colour blindness, etc.—were excluded, as well as those with any history suggestive of brain damage or any undergoing medication. Very few cases from amongst the trainee group had to be excluded on these grounds since their school selection had already screened out boys with major disabilities. However, a number were lost through apparatus failure. Finally, eighty subjects who were tested on all the procedures were obtained.

GENERAL PROCEDURE

Because the investigation being reported here was part of a larger whole, certain aspects of the overall procedure must be mentioned. The total testing time occupied six hours of one day, excluding appropriate resting periods. Two schedules of testing were devised which dovetailed so that two subjects were seen on each day. During the rest periods the subjects either read magazines or filled in the personality questionnaires. These latter comprised:

(i) the Maudsley Personality Inventory (MPI), yielding a score on Extraversion (E) and a score on Neuroticism (N)[1].

(ii) two forms of Cattel's contact Personality Inventory which seems to be measuring Extraversion (C) and a lie scale (D).

(iii) MMPI scales (Hathaway and McKinley, 1951), Hy, Hysteria; Pt, Psychaesthenia; Pd, Psychopathic deviance and K, the faking good, faking bad factor; these together seem to cover the areas of extraversion and neuroticism (Franks, Souief, and Maxwell 1959).

---

[1] Eysenck, H. J., *Manual of the Maudsley Personality Inventory*, Univ. of London Press, 1959.

The general intelligence test scores reported (test 90A of the National Institute of Industrial Psychology) were obtained from the apprentice training school records of the subjects.

Four techniques were used in the investigation being reported here. They were (*a*) Eyeblink Conditioning, (*b*) Salivary Conditioning, (*c*) Verbal 'Conditioning' and (*d*) Spatial 'Conditioning'. Half the subjects were put through the procedures in the morning and half in the afternoon. Each subject was seen by this investigator for about three and a half hours. The same order of administration was followed throughout the whole investigation. We shall deal with each procedure separately.

EYEBLINK CONDITIONING

Eyeblink Conditioning was included in this investigation on the basis of the previous work done by Franks (1956, 1957b) and by Das (1957). In his original study, Franks (1956) was able to discriminate a group of dysthymic patients (tense, anxious introverted neurotics) from a normal group and a group comprising hysterics and psychopaths. The hospitalized subjects were classified into groups according to twin criteria, a questionnaire measure and a rating by the relevant consultant psychiatrist on the basis of a check list of behavioural characteristics. When this rather careful selection procedure was carried out, it was felt that a fair degree of reliability in classification could be obtained and the hypothesis mentioned in the introduction could be tested. It was predicted that, if inhibitory tendencies predominate in the extraverts (the hysteric-psychopath group) then this group would condition less readily than the introverts (the tense, anxious dysthymics), in whom excitatory tendencies are held to predominate. This prediction was verified at an acceptable level of statistical significance and from this point Franks proceeded to test similar predictions in a variety of situations, including a confirmation of his finding with a normal group (Franks and Laverty, 1955, Franks, 1957, 1957a, Franks and Trouton, 1958). In general Franks' later studies have borne out his earlier findings.

Das (1957) was primarily concerned with the relationship between suggestibility, susceptibility to hypnosis and eyeblink conditioning. He also administered the Maudsley Personality Inventory, as did Franks in his study of normal subjects, and hence, the work of Das provides valuable comparison data with Franks' study. In fact, Das did not duplicate Franks' result and could report only a correlation of −0·083 between extraversion and conditioning whereas Franks had found a coefficient of −0·46 (Franks, 1957). This contradiction in results will be commented upon more fully in the later section in which the results of this investigation are discussed.

One other very important body of relevant work should be mentioned, that of Spence and Taylor and their associates (Spence and Taylor, 1951, 1952; Spence and Farber, 1953; Farber and Spence, 1953). Much of this work has been concerned with the relationship of 'Drive' to eyeblink conditioning where 'Drive' is present in the form of manifest anxiety, as measured by the Taylor Manifest Anxiety Scale (MAS). The researches of these workers derived from the theories and writings of Hull (1940, 1943, 1952) and of Pavlov (1927) very much as do those of Franks and that being now reported. The main difference between the two sets of studies is one of aim. Spence, Taylor and their associates are primarily and directly concerned with the generalized intervening variables of learning and motivation. Franks and the present writer, have been primarily concerned in following up the initial suggestions of Eysenck relating these variables to individual personality types.

If one accepts a Hullian account of learning and motivation as a working model then both drive, in the form of manifest anxiety (or as neuroticism, to use Eysenck's term) and inhibitory tendencies (associated with extraversion) should affect such a procedure as eyeblink condition, or any other learning or conditioning process. Spence and Taylor have claimed to have demonstrated the relationship of drive to eyeblink conditioning, but they, and those who have followed them, have largely ignored the at least equal importance of the inhibitory tendencies in such processes.

The specific aim of this part of the investigation then was to examine the relationship of the inhibitory tendencies and of drive, in the form of neuroticism, to eyeblink conditioning in a normal group of subjects. Eysenck and Franks have stressed the importance of the inhibitory tendencies. Our main emphasis will be on this line of reasoning and, following it, we would predict a negative correlation between a measure of extraversion and a measure of eyeblink conditioning. From the Spence-Hull position, one which, it should be stressed, is in no way in opposition to that of Eysenck, a positive correlation between eyeblink conditioning and a measure of neuroticism would also be predicted.

Apparatus

A detailed description of the procedure and apparatus for the eyeblink conditioning technique followed here has been given by Franks (1956) and by Franks and Withers (1955). Briefly, to summarize this information, the conditioning was carried out in a sound-proofed room, divided across its longer dimension by a curtain. In one half of the room the subject was seated in a chair provided with head and arm rests; in the other half, the experimenter sat before the apparatus for delivering the stimuli and recording the responses. The subject's field of vision was restricted by a three-sided booth, painted a neutral colour and having a small red light in the middle of the panel facing the subject.

The conditioned stimulus was a tone of 1000 c.p.s. delivered at 60 db. above each subject's bi-aural threshold, through a pair of high quality earphones, heavily padded to reduce ambient noise inside the sound-proofed room. The unconditioned stimulus was a puff of air delivered through a plastic tube at a pressure of 65 mm. of mercury. The two stimuli overlapped in time—the tone lasting 850 msec., the puff being delivered 450 msec. after the commencement of the tone. Both stimuli terminated together. Each stimulus could be delivered alone, or both be delivered together in the pre-arranged timing sequence.

The eyeblink response is recorded by means of a photoelectric cell held in the right hand lens carrier of a pair of spectacle frames which the subject wears. This carrier also holds the open end of the plastic tube through which the puff is administered. As an eyeblink causes the amount of light reflected from the eye to alter, a change in the electrical discharge from the cell is produced which, after suitable amplification, moves the ink writing pen of a recording milliameter.

Procedure

The subjects were told that certain automatic responses were to be measured. They were seated in the chair in front of the booth and the earphones and spectacle frames placed in position. At this time the subject's attention was drawn to the red light and they were instructed to watch it whenever it came on. This procedure aids in keeping the subject's head still during the crucial period within which the stimuli are administered. Any conditioned response to this light occurs before the onset of the tone conditional stimulus as the light is switched on four to eight seconds before the administration of the two crucial stimuli.

The subject's bi-aural threshold was then ascertained and the amplitude of the tone set to 60 db. above this threshold. Four conditioned stimuli alone were then administered to test for blinks to the tone as such. The subjects were then told that they would feel little puffs of air in their right eye. They were told to respond naturally to the puffs and to neither inhibit blinks nor to close their eyes to avoid them. The importance of watching the red light was stressed. Three puffs alone were then administered to accustom the subjects to this stimulus and to ensure that a blink could be elicited. Three more tones were administered to acquaint further the subject with the stimuli, he was asked if he were comfortable and if he understood what he had to do and the conditioning schedule was then commenced.

Eighteen conditional stimuli, dispersed amongst thirty reinforcement pairs in a partial reinforcement schedule comprised the first part of the conditioning procedure. The number of times a subject responded to the conditional stimulus when presented alone, constituted the principal score on the eyeblink conditioning procedure. As a secondary interest was whether anything analogous to reminiscence in learning occurred with conditioning, a second session was completed one hour after the first. Each session occupied half an hour.

Results

Table I summarizes the principal findings on eyeblink conditioning. It should be noted that Trials '1–18' refers to the first experimental session and Trials '19–37' to the second. The four questionnaire measures thought to be related to extraversion are listed first (Hy, Pd, E and C) and the four judged to be related to neuroticism are listed next (Pt, K, D, N). Of the correlation ratios in Table 3.1 only that between the two conditioning sessions is significant although that between N and trials

TABLE 3.1

Intercorrelations of the number of conditional eyeblink responses with various personality measures

	Trials '1–18'	Trials '19–37'	Hy	Pd	E
†32. trials '1–18'		+0·661**	−0·016	+0·038	−0·080
†33. trials '19–37'			+0·032	+0·018	−0·060

	C	Pt	K	D	N
†32. trials '1–18'	+0·028	+0·018	+0·126	+0·066	+0·183
†33. trials '19–37'	0·017	−0·141	+0·111	0·082	0·000

Henceforth ** will indicate a coefficient significant at the 1%
,, * ,, ,, ,, ,, ,, ,, ,, 5%
where there are 78 degrees of freedom
† Reference numbers of variables in principal component analysis.

TABLE 3.2

Comparison of mean number of conditioned responses in trials 1–18 *with that in trials* 19–37

	Mean	Standard Deviation
32. Trials 1–18	6·397	4·367
33. Trials 19–37	8·628	4·539

$n = 80$
$t = 5{\cdot}44$**

1–18 approaches very close to significance ($r = 0{\cdot}187$ is significant at the 5% level).

The analysis of the 'reminiscence' effect as such is to be presented separately by Franks and Willett (1959) but table 3.2 compares the mean number of conditioned responses in the first session with that in the second. Table 3.II indicates that a strong practice effect has operated from the first set of trials to the second.

Discussion

The results of this study are equally discouraging to an hypothesis linking extraversion via inhibitory tendencies to eyeblink conditioning or to an hypothesis linking anxiety-drive to eyeblink conditioning. The table below however shows that, with one exception, the present failure to find significant correlations between the measure of conditioning and personality measures is consistent with other work using the same procedure and apparatus, when normal subjects have been employed.

TABLE 3.3

Pearson product moment correlations between eyeblink conditioning and personality measures from five comparable studies using normal subjects

	'E' of MPI	'N' of MPI	'R' STDCR	MAS	MMQ
Franks (1954) N = 20 (male, female hospit. staff)			−0·048	−0·008	−0·145
Franks (1957) N = 55 (male students)	−0·46‡	0·04			
Das (1957) N = 68 (male students)	−0·083	−0·027			
O'Connor§ (1959) N = 14 (males)	−0·38				
Willett (Present study) N=80 (male apprentice trainees)	−0·08	0·183			

‡ Significant at 1% level of confidence.
§ N.B. O'Connor used the short version of the MPI (Eysenck, 1958).

The correlation coefficients, expressing the relationship between eyeblink conditioning and extraversion (R and E), shown in Table 3.3 may be regarded either as derived from five separate studies or as estimates of some true coefficient, obtained from five successive samples. Taken as derived from five separate studies, each coefficient is tested against zero. On this view, it can be seen that the result obtained in this study is consistent with three of the other four studies. Considered as derived from five successive samples varying about some true value, the five coefficients would seem to be varying about a relatively small, negative value. That this value is unlikely to be zero is indicated by the fact that all five coefficients are negative. The negative sign is, of course, in accord with the predicted relationship between eyeblink conditioning and extraversion. Furthermore it would seem reasonable to suppose that Franks' finding of a coefficient of $-0{\cdot}46$ (1957) differs markedly from the others, as no other coefficient could be distinguished from zero while it differs significantly from zero.

To obtain an estimate of the true coefficient a z transformation of the coefficients was made (Fisher, 1948) and an average weighted z score computed. This z score was transformed back into product moment coefficient form and tested against zero. The estimate of the true coefficient obtained by this procedure was $-0{\cdot}188$ which, for 227 degrees of freedom ($\Sigma(n - 2)$) is significant at the 5% level.

That Frank's' result (1957) seems the exceptional case has been commented upon. If we are dealing with successive comparable samples it is difficult to see why this should be. The same conditioning procedure was used in all studies and comparable questionnaires were administered; Franks (1954) used the 'R' scale of the STDCR, from which the E scale was derived and O'Connor used a shortened version of the MPI. In the other three studies the full version of the MPI was employed. Where any differences in questionnaires do exist, in short, no differences are evident between coefficients; Franks (1957) used the full version of the MPI. Some differences between the subjects of the studies are apparent and this could be important. Here again, however, although Franks' (1957) subjects were different from those of his earlier study (1954), from those of O'Connor (1959) and from those of this study, they came from the same population as did those of Das (1957) whose results are comparable with all others except Franks' (1957). Franks' (1954) himself considered the importance of age, sex and intelligence to conditioning and concluded that these factors had relatively little bearing on this process except at extremes of age and intelligence. If some important variable was controlled in the selection of Franks' (1957) subjects and not in the selection of the other subjects, it is not obvious from his report what this might have been and what its influence on the other studies could have been. We cannot ignore the possibility that environment factors, whose influence on conditioning is as yet unknown, may have an important bearing on the formation of conditional responses. In the absence of any such knowledge, however, Franks' finding appears as a sampling deviation and we must conclude that a small negative correlation represents both the most frequent finding and the best estimate of the true correlation between extraversion and conditioning.

One then feels constrained to ask whether this represents a fair test of the prediction derived from the basic hypothesis relating eyeblink conditioning to extraversion and, immediately, the possibility that the questionnaire measure being employed is a poor criterion of this parameter demands scrutiny. To state this possibility succinctly; we are concerned with showing that conditioning is related to extraversion in a particular way because both are related to inhibitory tendencies. Where normal subjects have been employed it is conceivable that the questionnaires used do not adequately differentiate between the groups in the manner required to test the hypothesis.

No direct evidence can be brought to bear in support of this suggestion, however, the following should be borne in mind. When he obtained significant coefficients with his abnormal groups, Franks (1954) used a double criterion, as has been pointed out. The questionnaires were administered, as in the later studies, but the obtained scores had to agree with ratings on clearly defined behavioural characteristics made by the psychiatric staff before a subject was included in one of the tested groups. Thus up to a point, behaviour ratings were guiding the selection and it would happen then, even if the questionnaire was functioning inadequately, something approaching a satisfactory differentiation could have been achieved by virtue of these ratings. An ongoing investigation by Claridge throws further light on this problem. He administered a group of tests to psychiatric subjects where, as in eyeblink conditioning, certain predictions were made connecting performance to personality measures. He found that, in many cases, no relationship could be found when questionnaire measures were used but that employing the same kind of behaviour ratings as Franks employed, the relationships, as predicted, did appear. Sigal, Star and Franks (1958), investigating the ability of the MPI to discriminate between a group of dysthymics and a group of hysterics selected on case history symptom ratings by three raters, showed that no differentiation on the basis of the questionnaire scores could be made between these two groups.

Furthermore, if the questionnaire measures are not involved, predictions based on the central hypothesis fare reasonably well. Das, for example, found the predicted correlation between hypnosis and conditioning. There is a reasonable level of consistency in the drug studies reported here and elsewhere (*vide* Volume One) and a certain degree of order can be brought to the behavioural effects of psychosurgery (Willett, 1959) to mention only some of the instances where the concept has proved useful.

However, there is no gainsaying the fact, dramatically illustrated in Table III, that when predictions involve questionnaire measures they rather too frequently fail to be verified. Note also for example, the almost total absence of any significant correlations in the section on drug effects despite the appearance of many predicted effects. One is thus inclined to conclude that when the Maudsley Personality Inventory is used (to take our principal questionnaire) to give criterion scores, it gives very poor ones. So poor, that no information is added to that which might be obtained from investigating the processes themselves.

This writer feels that progress in describing the connection between personality parameters and fundamental explanatory processes is being retarded by the lack of a more valid personality measure. The negative results of this investigation exemplifies this lack.

SALIVARY CONDITIONING

A procedure hallowed by its associations with the animal investigations of the early Russian physiologists, Salivary Conditioning has been frequently studied in human subjects. (Gantt, 1936, 1938; Razran, 1933, 1935, 1936, 1939, 1955a, 1955b; Finesinger and Sutherland, 1942; Bindra, Paterson and Strezelecki, 1955.) Since this procedure seemed to have been employed successfully by these workers and because of its connections with classical conditioning experiments it was decided to include salivary conditioning in this investigation. As before, with eyeblink conditioning we were concerned here with the relationship of salivary conditioning to the personality dimensions of introversion/extraversion and neuroticism as measured by the various questionnaires already described. It was expected that a negative correlation between introversion/extraversion and salivary conditioning would be found. The salivary conditioning procedure was tried out in the pilot study and satisfactory correlations were found between a score on this function and E and N of the MPI. (r(E and Saliv. Condit.) $= -0{\cdot}57$, r(N and saliv. condit.) $= 0{\cdot}54$), both being significant at the 5% level of confidence).

This finding was sufficiently encouraging to warrant proceeding with the major study to be described. The procedure in both the pilot study and the main study were the same.

Procedure

An adaptation of the procedure described by Razran (1936) was followed, particular attention being paid to his observation that subjective attitudes affect the formation of salivary conditioned responses in human subjects. The subjects were therefore required to practise on a ball-rolling maze game which had been previously shown to hold the interest of males, over a period of at least two hours.

The task consisted of a maze path described between sets of numbered holes on a tray, which could be tilted, either up or down or from left to right, by manipulating two knobs. The subject was instructed to roll the ball along the track as far as he could by tilting the tray, without allowing it to fall into one of the holes. He was told that his score was the furthest point he reached, that it was an advantage to proceed carefully and slowly, that he was not being timed and that if, at the end of a trial the ball was still on the tray, he was to leave it where it was and commence from that point when the next trial began. He was allowed a short time to practise manipulating the knobs and the experimenter demonstrated that it was possible to roll the ball a considerable distance along the track without committing an error.

Technique of Conditioning

Fitted to the maze game and elevated three inches above it were a red lamp and a white lamp each of 6 volts. These could be switched on or off independently by the experimenter and provision was made for the white lamp to flash on and off once a second.

Razran's technique (1935) of measuring increments in salivation by the insertion of cotton rolls under the tongue was followed (Johnson and Johnson, No. 1), with the modification that, on receiving the relevant signal, the subject placed the roll under his tongue *himself* with a pair of sterilized surgical tweezers and removed it in the same way instead of the experimenter carrying out these operations.

The experimental session occupied about an hour and forty minutes and consisted of a number of 'periods' as follows.

(1) The reinforcement or training periods consisted of the subject practising on the ball maze with the red light switched on and the white light flashing (CS), while sucking 25 mgm. tablets of l-ascorbic acid (USC). There were two reinforcement periods of ten trials, each trial lasting for two minutes and separated by rest pauses of thirty seconds.

(2) Periods of 'regular control' consisted of the subject practising on the ball maze with both lights switched off. These consisted of 45-second trials and one trial followed each of the reinforcement periods to ensure that the saliva secreted during the latter was absorbed.

(3) 'Test' periods consisted of the subject practising on the ball maze with the red light on and the white light flashing with the cotton rolls in position. These consisted of 45-second trials, three trials following the first reinforcement period and five following the second.

(4) 'Control' periods consisted of the subject practising on the ball maze with both lights switched off. These consisted of 45-second trials and one was paired with each 'test' trial, so that three followed the first of the reinforcement periods, and five followed the second.

'Test' trials and 'control' trials thus came in pairs and the order of presenting 'test' trials and 'control' trials was alternated throughout the eight pairs and from subject to subject.

After the subject received his instructions concerning the ball maze, he was given the following instructions relating to the cotton rolls and the signals he was to act upon.

'While you are practising on this game I am going to take a number of measurements from you that will require you to place these rolls of cotton underneath your tongue. You will use these tweezers to place the rolls under your tongue and also to remove them. You will place and remove the rolls, and commence and stop practising the game to certain signals. The signals are as follows.

A. (Relevant to the 'test' trials.) 'When the red light comes on you are to place the rolls under your tongue. When the white light begins to flash, commence to practise on the game immediately. When the lights go out you are to stop practising on the game immediately and remove the cotton roll from under your tongue and hand the roll to me, in the tweezers.

B. (Given prior to each of the reinforcement periods.) 'This time when the red light comes on I want you to take one of these tablets, place it in your mouth and commence practising on the maze when the white light begins to flash. I don't want you to chew the tablet or swallow it—it won't hurt you if you do because it is only a flavouring matter; it tastes like a sweet—but I want you to suck it all the time you are practising on the maze. When the lights go out, stop playing the game immediately and if there is anything left of the tablet, remove it and place it in this basket.

C. (Relevant to the control periods.) 'When I say "now" you are to place a roll of cotton wool under your tongue and commence to practise on the game immediately. When I say "stop" you stop playing the game and hand me the roll in the tweezers.'

Summary of the procedure

(*a*) Instructions to S concerning game and cotton rolls.

(*b*) Three 'introductory' trials consisting of three pairs of 'tests' and 'control' procedures, each of 45-second duration.

(*c*) Ten reinforcement trials with the red light on, the white light flashing and S sucking tablets of 25 mgm. l-ascorbic acid; each trial of 2 minutes duration, with 30 seconds rest between trials.

(*d*) One 'regular' control trial of 45 seconds duration.

(*e*) Three pairs of 'test' trials and 'control' trials each of 45 seconds, the order of 'test' and 'control' being varied throughout the three pairs and from S to S.

(*f*) Ten more reinforcements as in (*c*).

(*g*) One 'regular' control as in (*d*).
(*h*) Five more pairs of 'test' and 'control' trials as in (*d*).

In periods '*b*', '*d*', '*e*', '*g*' and '*h*', the rolls were weighed immediately on removal from S's mouth by means of a torsion balance. They were also weighed before being placed in position, to allow for slight variations in the size of the rolls.

Results

Table 3.4 shows the correlation coefficients between the two related measures of salivary conditioning and the questionnaire measures. In this table $\Sigma\frac{(T-C)}{n}$ is the mean difference between the 'test' trials and each paired 'control' trial. Other things being equal, the larger this difference the greater the degree of conditioning achieved. Because such very large variations occurred within and between subjects in the amount of salivation, a second score was taken from the simple count of the number of times the salivation on the 'test' trials exceeded the salivation in the paired 'control' trials. This score we have called 'No. of +'.

TABLE 3.4

Pearson product moment correlations between two measures of salivary conditioning and various personality measures

	$\Sigma\frac{(T-C)}{n}$	No. of +	HY	Pd	E of MPI
34. $\Sigma\frac{(T-C)}{n}$		+0·712**	−0·114	+0·173	+0·095
35. No. of +			−0·116	+0·018	+0·077

	'c' of CPF	Pt	K	D	N
34. $\Sigma\frac{(T-C)}{n}$	+0·078	+0·122	−0·319**	−0·134	+0·079
35. No. of +	−0·067	+0·193*	−0·349**	−0·222*	+0·142

Apart from the understandably high intercorrelation between the two conditioning measures, four reasonably low, but significant correlations with measures of neuroticism are evident in Table 3.4.

The individual scores on this conditioning measure were characteristically extremely variable, particularly from one subject to another but also even within one subject's record. So much was this the case that one could not tell just by looking through the array of scores whether on the whole salivation was greater during the 'test' periods than during the paired 'control' periods. This feature of the scores is well illustrated by stating that the mean of all the $\Sigma \frac{(T - C)}{n}$ scores was only $+ 14{\cdot}79$ while the standard deviation of these scores was 65·38. It should be remembered that $\Sigma \frac{(T - C)}{n}$ is the mean of the differences between the eight pairs of 'test' and 'control' scores for any individual and can be either positive or negative.

To ascertain whether the mean of the $\Sigma (\frac{T - C)}{n}$ did differ significantly from zero (note that the mean is positive indicating a higher degree of salivation under the 'test' condition), a simple '*t*' test was carried out using the values for the mean and standard deviation already given. For the eighty cases of our study this computation yielded a '*t*' value of 2·00 which, for this number of cases indicated that the mean of the $\Sigma \frac{(T - C)}{n}$ scores did differ significantly from zero ($P = 0{\cdot}023$), a requirement to be fulfilled if conditioning is being claimed to have occurred.

It still seemed possible that although the 'test' trials were producing a significantly greater amount of salivation this was due, not to any conditioning process but to some simple effect of the blinking lights used as the conditioned stimulus. To test this possibility a control study was carried out on a sample of twenty cases drawn from the same parent population as that of the main study. These subjects were put through exactly the same procedure as in the main study with the exception that during the twenty two-minute reinforcing periods no 1-ascorbic acid was administered. In other words no reinforcing situation was present and the comparison that could now be made between the 'test' and the 'control' periods was simply the effect of the blinking lights present in precisely the same context as in the main study. This time the mean difference between the 'test' and 'control' measures of salivation, computed exactly as before, was +29·4 and the standard deviation of the array of mean differences was 74·83. Again the significance of this value's departure from zero was computed and the '*t*' value this time equalled 1·7 which, for the twenty cases used here, indicates that the departure from zero of this mean cannot be viewed with as much confidence as that of the main study ($P = 0{\cdot}053$).

The scores of both the main study and the control study were examined to discover whether any increase in amount of salivation throughout the period to testing occurred but in both cases the sum of the first two measurements taken over all the subjects was almost precisely the same as that of the last two.

Finally the correlation between amount of salivation and E and N was computed. The salivation score used here was the sum of the six 45-second trials (three 'test' trials and three 'control' trials), given at the commencement of the procedure to accustom the subject to the technique of inserting and removing the cotton rolls. A correlation coefficient of —0·13 was obtained between amount of salivation and extraversion (E) and of 0·13 between amount of salivation and neuroticism (N). Both personality measures were taken from the Maudsley Personality Inventory and neither value was statistically significant.

Discussion

The degree of variability present in the salivary responses in this investigation seems a most unsatisfactory feature of this technique to this writer. As has been pointed out, the variability was not only between subjects but within the responses given by any one subject. The whole range of the torsion balance was used, some subjects giving readings of less than 100 milligrams and others ten times as much. Individual variation was not quite as great but ratios of three or four to one from one response to the next were not uncommon. It was the noting of this variability in the pilot study that led to the practise of alternating the order of 'test' and 'control' trials within and between subjects and, in consequence of it, one cannot feel very confident that conditioning was achieved at all. The control study did clarify this situation somewhat and suggested that the conditioning apparently achieved was an artifact of the stimuli used. Thus the use of the salivary conditioning measures to further our knowledge of the relationship between personality parameters and conditioning must be viewed in the light of the uncertainty concerning the measures employed.

It was recognized early in the investigation that, to a far greater degree than is the case in eyeblink conditioning, the stimuli of salivary conditioning are very complex. The conditioned stimulus in our procedure is clearly not only the blinking lights but the whole complex of placing cotton rolls in position, playing a ball game, being timed, seeing cotton rolls being weighed and so on. Almost all of this complex is carried over into the control situation, the only difference being the change of signals and the absence of the lights. Certainly, as the experiment was carried out in a room only dimly lit, the absence of the lights was very conspicuous, nevertheless, the degree of similarity between the 'test' and 'control' situations was considerable. It was for this reason that the scores were examined for any overall increase in amount of salivation throughout the course of the testing session. A general rise in salivation could have been a reflection of a conditioned response of salivation developing to the total stimulus complex. The control study proved unnecessary from this point of view since no overall increase was evident in it or in the main study.

We examined the relationship of amount of salivation to the personality parameters as an interesting side aspect of the investigation because, salivation being closely related to antonomic activity, it was thought that a relationship might exist between this measure and neuroticism. Wenger (1948) included salivation as one of the indicators of antonomic activity in his study of antonomic balance in army air-force personnel. In a series of comparisons between aviation students, psychiatric groups and 'Operation Fatigue' groups it would appear, from Wenger's study, that with some inconsistencies, salivation discriminates between groups where the degree of emotional stress varies. In general, the direction of the difference is opposite to that one might have expected. In all but one of the comparisons where significant differences were found, the group characterized by being rather more ill showed less salivation. In a centroid factor analysis of his data Wenger identifies salivation as an indicator of antonomic activity although, in fact, its loading on the factor he identified with antonomic activity was only 0·25.

Wenger also correlated his physiological measures with the scales of the STDCR. In agreement with the results reported in this study he found no significant correlation with D and C (elsewhere shown to be clearly related to neuroticism, Franks, 1954) but reports a significant but very low correlation of +0·18 with R from which the E scale of the MPI was derived.

Coming to the principal result then we shall have to bear in mind the points already made. Namely, that whether conditioning occurred at all, seems problematical and salivation, as such, does not seem to be very closely related to either neuroticism or extraversion, as measured by the Maudsley Personality Inventory or the D, C and R scales of the STDCR.

We note first of all that, as in eyeblink conditioning neither E nor N of the Maudsley Personality Inventory shows any sizeable relationship with the measures of salivary conditioning. The four significant coefficients found (apart from that between the two conditioning measures) are with other measures only associated with neuroticism. Those with K, the 'Faking good, Faking bad', scale of the MMPI, are significant at beyond the 1% level.

D here is the lie scale of the Contact Personality Factor Inventory and gives the next highest correlation with the salivary conditioning measure. If in fact these significant correlations have not simply been thrown up by chance from amongst the very many computed, does this result mean that these scales are better measures of personality dimensions relevant to conditioning than the other scales? It would seem not, for neither show any degree of concomitance with eyeblink conditioning, a more reliable conditioning procedure than this of salivary conditioning.

The correlations, in the case of both K and D, are negative such that high scores on these questionnaire measures are associated with low scores on salivary conditioning. In this study both K and D have negligible correlations with extraversion (E of MPI) but both have highly significant negative correlations with N (of MPI) such that high scores on K and D are associated with low scores on N. It should also be noted that K and D correlate with each other to the extent of 0·5.

If anything is to be made of these facts it must be that persons with low scores on neuroticism also get low scores on salivary conditioning. In other words since K and D are in a sense the inverse of N and K and D correlate negatively with salivary conditioning, a positive relationship between salivary conditioning and N might be said to have been demonstrated. However, neither K nor D functions in the same way in relation to the other measures of conditioning, as has been pointed out.

Bearing in mind the connection between N and D (of STDCR) and Wenger's (1948) results, the correlation between K and amount of salivation was computed in the same way as was done with E and N. Again this measure of salivation bore no relationship to the personality measure ($r = 0{\cdot}08$).

In general one must conclude, as in the case of eyeblink conditioning, that no meaningful or consistent relationships between this conditioning measure and the personality parameters has been demonstrated. The criticisms one made in the section on eyeblink conditioning about the inadequacy of the personality measures applies in this case also of course, but with this measure the situation is worsened by the uncertainty of whether conditioning occurred at all. No firm conclusion either negative or positive can be drawn from the results of this study therefore.

VERBAL CONDITIONING

The commencement of interest in the phenomenon of Verbal Conditioning, or Probability Learning, as it has more appropriately come to be named, is commonly associated with a study of Humphrey's (1939). In this study the subject was required to say whether, when a light came on, a second light would follow it or not. In Humphrey's study the second light followed 50% of the time. Grant, Hake and Hornseth (1951) employed ratios of 25%, 50% and 75% in repeating Humphrey's study. In both cases it was found that subjects would match the frequency of their 'yes' guesses about the second light to the actual probability of its appearance, although the sequence of appearance and non-appearance was, to all intents and purposes, random. This phenomenon of frequency matching we have investigated in its relationship to individual personality measurements by examining the rate and regularity with which subjects changed from one response frequency to another.

We have taken the simplest case of probability learning where cues are minimal. The determining cognitive conditions and general ramifications of this and of related processes are many. An excellent account of these will be found in Bruner, Goodenow and Austin (1956).

The general design of the experiment required the subject to accept first a particular frequency of appearance for one of a pair of stimuli in an initial series of trials. The ratio chosen was that used by Humphrey, 0·50. In other words on half the choice occasions one choice was 'correct' and in half, the other. It was hoped thus to establish a frequency matching tendency in this period. Then, without the subject's knowledge, the ratio was changed so that, approximately one choice was 'correct' twice as frequently as the other. It was hoped to plot the course of adapting to the new frequency ratio and to relate, particularly, the rate of change to the personality variables.

If a conditioning model is appropriate to this task (a matter to be discussed) then one would expect from the hypothesis about extraversion and conditioning that subjects with low scores on this factor (i.e. the introverts) would readily match the stimulus frequency in their responses and then only slowly change to the new frequency. That is, they would 'condition' well and extinguish slowly. The persons with high scores on extraversion would not match the initial frequency as well and would change more easily but more erratically. In the pilot study a positive correlation was obtained between rate of change and extraversion and a negative correlation between variability and extraversion. Neither was significant statistically but they were of sufficient size to warrant the work undertaken in the main study.

Apparatus

The apparatus consisted of two white six-volt lamps set eighteen inches apart in the middle plane of a grey board three feet by two feet. Nine inches above the mid-point of the line joining the two white lights was a six-volt red lamp. The white lamps were designated left (L) and right (R) and the red lamp referred to as the 'signal light'. The board was held in a stand so that these three lights faced the subject.

Behind the board, where the experimenter sat, were a duplicate set of lights, which came on simultaneously with the lights seen by the subject. This was to ensure that, barring bulb failures, the experimenter would be immediately aware if any of the lamps failed to work. In the rear base of the board the experimenter was provided with two switches. One controlled whether any of the three lights would come on and the other, a selector switch, allowed the experimenter, once the first was closed, to select which of the two stimulus lights would come on. Care was taken to ensure that both switches were silent. The selector switch was interposed between the lights and a timer. The timer was arranged so that, with the switches in the appropriate position, the red light would come on for five seconds, then one, the other or both of the white stimulus lights would flash automatically. The experimenter was required merely to select which stimulus light was to follow the signal light. The timing, selected on the basis of the pilot study, was: Signal light on for five seconds, stimulus light flashed (≏ $\frac{1}{5}$th second), no light on for ten seconds, signal light on, etc.

Procedure

The subject was seated before a desk in a curtained room facing the front of the apparatus. He was instructed in the following terms:

'You will notice that on this board there are two white lights marked left and right and a red signal light. Every fifteen seconds or so the red light will come on and after a few more seconds one of the two white lights will flash like this . . . (the subject was then shown one or two cycles to acquaint him with the stimuli). Now I want you to tell me each time the red light comes on, which one of the two white lights is going to follow it. You must say "left" or "right" before the white light flashes for your answer to count and, as you have seen, you have not got much time.' (The subject was then given a few trials to accustom him to the task of responding.)

After any further explanations that were required the investigation began.

The experimenter was provided with record sheets on which the two schedules were laid out with spaces opposite for recording the subject's responses.

Preliminary testing had shown that, as was suggested in the literature (Grant, Hake and Hornseth, 1951), seventy trials were adequate to establish a frequency matching of 50% and consequently a schedule of seventy trials, in which equal numbers of left and right choices were 'correct', was arranged by reference to a table of random numbers. Likewise, preliminary testing had indicated that a ratio of seven to thirteen was sufficiently 'difficult' to yield a good range of matching rates through seven sets of twenty trials. Accordingly this ratio and this number of trials was adopted in the main study. Within successive blocks of twenty trials, then, in the second series of trials, seven left 'correct' choices and thirteen right 'correct' choices were arranged by reference to a table of random numbers.

It was found that very few subjects failed to respond in the choice situation. If this happened more than four times the test was abandoned. In those few cases where failures occurred, other than these, the subject's record was marked with the 'correct' choice. The number of occasions when this occurred was, however, negligible.

Results

That frequency matching occurred in conformity with the first schedule of seventy trials, where the ratio of left to right correct choices was 0·50, is shown by the mean number of left choices made by the eighty subjects. This mean was 34·69 ± 3·80. Perfect matching would have been indicated by a mean of 35·00 ± 0 left choices.

That adaptation to the new ratio did occur to an appreciable extent is shown by Table 3.5.

TABLE 3.5

Mean number of 'left' choices made through seven series of twenty trials where the ratio of left to right correct choices became 7 *to* 13 *after seventy trials where the ratio was* 0·5

	Successive sets of twenty trials 1.	2.	3.	4.	5.	6.	7.
Mean no. of Left choices	11·84	9·61	10·25	9·90	8·20	8·61	8·70

Had a perfect and smooth adaptation occurred in this group the mean values would have dropped from rather less than a mean of 10·00 in the first block of twenty trials to one of 7·00 in the seventh block.

To examine whether subjects with low scores on extraversion gave closer and more consistent frequency matches than subjects with high scores on extraversion, the mean number of left choices and the appropriate standard deviations from the first schedule of seventy trials were computed for the group whose E scores were below the mean E score of the whole group and for the group whose E scores were above the mean E score. The mean and standard deviation of the number of left choices for the low E scorers (the introverts) was 34·12 ± 3·32 and that for the high E scorers, 35·11 ± 4·10. These values are virtually indistinguishable.

In examining the main predictions associating conditioning with personality parameters three scores were computed: a score which reflected the gross change in the number of left choices made by the subject, a score which reflected the rate of change and a score which indicated the variability of the number of the subjects 'left' choices over the seven blocks of twenty trials. Since there were half the number of trials in the first schedule of trials as in the second, the first score was obtained simply by subtracting twice the number of left choices in the first schedule from the total number of left choices in the second schedule of one hundred and forty trials.

TABLE 3.6

Pearson product moment correlations of, gross change in left choices from the first schedule of 0·50 *ratio to one of* 7:13, *rate of change and variability of number of left choices over seven sets of twenty trials with personality measures*

	Change in Left Choices	Rate of Change	Vari-ability	HY	Pd	E of MPI	C of CPF	Pt	K	D	N
42. Change in Left choices		+0·219*	−0·236*	−0·182	+0·008	+0·077	−0·052	+0·241*	−0·120	−0·008	+0·221*
43. Rate of Change			−0·025	−0·093	−0·122	−0·118	+0·051	−0·109	+0·117	+0·180	−0·186
44. Variability				+0·218*	+0·220*	+0·083	0·044	−0·025	0·029	0·078	−0·157

The scores for rate of change and variability were obtained in the following way. The number of left choices in each of the seven blocks of twenty trials, for each subject, were treated as points lying about a straight line. This straight line of best fit was computed for each subject and its slope determined. This slope then represents rate of change. Variability was expressed as the sum of the squared deviations of the plotted points from the corresponding points on the straight line of best fit.

The direct measure of 'conditioning' here is the rate of change measure, although some interest attachès to the measure of variability and to the gross change scored. Table 3.6 shows the correlation of these measures with the personality parameters.

In the above table Pt and N correlate significantly with the gross change in number of left choices while HY and Pd correlate significantly with variability.

Discussion

Unlike the case of salivary conditioning there seems little doubt that here the procedure as such was reasonably successful. A frequency-matching response that was very close to perfect was established in the group as a whole and the extent to which an individual's scores varied about the mean matched frequency was quite small. With the introduction of the second schedule and the new 'correct' stimulus to 'incorrect' stimulus ratio, a gradual adaptation in the responses did occur towards an approximation to the new ratio. The decrease in the number of 'left' choices made by the subjects was small but consistent. Only 15% of the subjects did not give responses yielding a negative, i.e. falling, slope. That the gross change score was also negative is a reflection of the same fact.

When one turns to the relationship between these scores and the personality measures, however, one again has little that is positive or conclusive to report. If the extraversion scale of the MPI is to be judged as a valid measure of this personality parameter then there was no indication in this study that introverts returned more closely matched frequency responses than extraverts.

Some positive correlations are present between the scores derived from the second schedule and the personality measurements. Neuroticism, from the MPI and psychaesthenia from the MMPI both correlate positively and significantly with the change scores. Likewise the hysteria and psychopathic deviate scales of the MMPI correlate positively and significantly with variability.

Taking the case of neuroticism first, it should be noted that the correlation between N of the MPI and Pt, in this study was + 0·538. Thus, although with a correlation of this size the variance accounted for by the concomitance is quite small, we have some justification for regarding the two scales as measuring the same thing. If we call the variable being measure by both scales, neuroticism, for the purposes of this discussion and bear in mind the size of the correlation between the two personality scales, we will not then do the data too much violence.

Each scale correlates with the gross change score to almost the same extent, 0·241 in the case of Pt and 0·221 in the case of N. Because of the way in which the change score is derived (and because the array of change scores is very largely negative) the larger the change the larger *negative* quantity the change score will be. Thus if the relationship between change score and neuroticism were truly positive, the correlation coefficient would have appeared negative. Our coefficients mean then that the smaller the change the greater was the degree of neuroticism. This result makes little sense on a Hullian-Spence view of neuroticism nor in the terms in which Eysenck might reformulate their view. The curvilinear relationship between performance drive and task complexity could account for the result but unless one can, by varying experimental conditions, demonstrate the law operating inversely the *ad hoc* involving of this effect is not very convincing. An *ad hoc* explanation which, however, does seem to have greater face validity is that the two personality scales are, in part, measuring a rigidity factor and that the change score represents a rigidity score such that now the coefficients are to be understood as indicating: the smaller the change (the larger the performance rigidity) the greater the (questionnaire measured) rigidity.

That the rate of change score does not reflect the same tendency renders this suggestion a little more tenuous than if the results had been consistent but, as the correlation between rate and gross change is only 0·219, the proportion of the variance this coefficient explains is very small indeed and there is a great deal of room for other factors, including error, to play a part. The two correlations with rate are, in fact, insignificant, as are all the other coefficients involving the rate score. This rate of change, it was initially felt, would best indicate the conditioning procedure, if the procedure can be regarded as conditioning at all.

Variability yielded two significant correlations with the HY and Pd scales of the MMPI. These scales correlate together significantly (+0·205) but neither shows any large degree of concomitance with either E or N of the MPI. Variability was not regarded as a conditioning measure but it was expected that a more variable performance might be expected of a subject who was not adopting any particular frequency-matching ratio. Again, however, we have no evidence that our expectation was borne out. Unlike the study on the basis of which the questionnaire scales were grouped (Franks, Souief and Maxwell, 1959), HY and Pd do not correlate, in this study, with extraversion, and extraversion does not correlate with the performance measures. There is no evidence to suggest even that these two scales are better measures of extraversion than E since the two scales yield no other comparable correlations with any of the other performance measures. In fact one cannot avoid concluding that these ratios seem more than likely to have been thrown up by chance.

What kind of learning process does probability learning most closely resemble, one must now ask? That an adaptation and modification of the subject's responses occurs is beyond doubt. That this is not primarily a question of intelligent problem-solving behaviour, in which insight plays an important role, is largely ruled out by the nature of the task and contra-indicated by the lack of any sizeable or significant correlations between scores on the procedure and scores on the test of general intelligence (NIIP test 90A: r_{change} and $90\text{A} = -0{\cdot}184$ r_{rate} and $90\text{A} = -0{\cdot}029$, $r_{\text{variability}}$ and $90\text{A} = 0{\cdot}153$). This observation does not imply that the behaviour does not 'occur under a problem-solving set' (Bruner, Goodenow and Austin, 1956) but serves to emphasize that if probability learning is to be regarded as a problem-solving process then it is more analogous to trial and error learning or to some other progressive process than to insightful solution finding. The special properties of insight are an awareness of and a suddenness in solution and a general reperception of all the elements of the problem in the new light of the solution.

Now this is not the situation we have at all in the procedure we have used. As we have pointed out the progression towards correct 'solution' is gradual and the subject is not aware that he has reached it. The correctness or otherwise of his solution is only statistically implicit in his responses. Far from seeing the sequence of choice points in some new light whose source was a correct solution, very often subjects were convinced that they had discovered the 'pattern' of lefts and rights which the experimenter was following when, in fact, no pattern existed, other than the purely statistical one of an over-all frequency of occurrence for the 'left' light.

Trial and error learning involves, essentially, the selection, by some form of reinforcement, of one response amongst a number. Now in the probability learning situation a maximum score can only be achieved if the subject matches the choice perfectly. A perfect match doesn't guarantee a maximum score but it is the only response pattern that can give one and the subject has been instructed to attempt the maximum score. A detailed analysis of the parameters involved in this process is to be found in Bruner, Goodenow and Austin (1956) and concerning the simple type of probability learning employed in this investigation they describe subjects as being guided in their choices by the overall frequencies and also by certain other specific factors. The two mentioned as most important are the tendency to regard events associated in time as connected in other ways (i.e. of being mutually dependent) and the tendency to 'operate on the basis of shorter runs of events than occur on a choice basis'. Now it seems to us that the subject's progressive approximation to a true frequency may most fruitfully be viewed as a series of trials, necessarily continuous through the nature of the problem, in which certain responses (incorrect matches) are dropped out until, by the touchstone of the recent run of successes and failures, an approximation to the correct response (i.e. the true frequency match) is made, tested and adopted as the 'correct' maximally rewarding response.

Trial and error learning does seem to be a reasonably good model for this process, then. Can it be regarded as conditioning in any sense? After experience with the procedure it seems doubtful. The conditioning paradigm is a simple one. When one stimulus is associated with another which of its nature elicits a response, the response to it becomes attached to the associated stimulus which previously did not elicit this response. What in probability learning would we regard as the unconditional stimulus? The response to be conditioned is the true frequency. Can this frequency, in the stimulus schedule, be said to elicit the matching response in the same way as, for example, the puff of air elicits an eyeblink in the eyeblink conditioning procedure? It would seem that it cannot and if the matching response were so regarded what would one regard as the conditioned stimulus? Could the lights themselves be regarded as the conditioned stimulus? No, because within the terms of the investigation the lights can never be in any context, other than that of some frequency of occurrence ratio to which subjects will adapt. In other words the conditioned stimulus is inseparable from the stimulus which elicits the unconditional, to-be-conditioned response.

The best way of viewing the procedure, from the point of view of the general aims of the investigation then seems to be, not as a true conditioning process but as an associative or learning process in which a response tendency is built up and modified to meet a changed condition. It would not have been surprising, if had positive results been reported in the cases of the two previously described investigations, this technique would have produced a different result. Since all three investigations were inconclusive one can hardly compare them. On the whole, however, one would have expected the extravert group to be relatively less efficient in their learning here also and hence one could, quite legitimately, investigate personality differences in this type of process and, if the data were at hand, could quite justifiably look at the various interrelationships between the processes and the personality parameters. The results that were in fact obtained, however, are too fragmentary to base any conclusion upon even if the process is regarded in this way and not as true conditioning.

SPATIAL 'CONDITIONING'

The technique employed in this investigation was suggested by work carried out in the study of synaesthesia.[1] The phenomena of synaesthesia refers to the association of the characteristic response of one sense modality with a stimulus usually relevant to another modality. This study has led to an examination of how one type of stimulus (say a tone) affects the perception of another. That an effect of this type is produced[2] suggested the procedure to be described. Essentially, this procedure attempted to cause subjects to associate the location of a tone in a semi-circular horizontal plane with the position of a light in the same plane.

Again, the procedure was initially conceived as an example of conditioning and it was expected that measures on the performance task would be related to personality parameters estimated from questionnaires. More particularly it was expected that any association would be less marked for the subjects with high scores on extraversion.

Apparatus

The apparatus consisted of a semi-cylindrical booth whose long axis was in the vertical plane. The booth was constructed of metal supports joining two semi-circular wooden members which formed the top and the base of the booth. The whole framework was covered with a white blanket material which was reinforced on the top by heavy felt. A scale marked on a strip of heavy drawing paper eight inches wide was glued horizontally around the inside of the booth so as to be at eye level for a sitting subject. A second and identical scale was fixed behind the first for the use of the experimenter. The scale was marked with zero at the centre and with fifteen divisions by fives, to the left and to the right. The positions to one side were marked in green and those to the other in red so that one half of the scale read from zero at the centre to 75 green at the periphery and the other read from zero to 75 red at the periphery. At each end of the scale was a 'pearl' 25-volt lamp enclosed in a shade which gave a diffused light of low intensity inside the booth.

Two arms were fixed at the locus of the top semi-circular member, one above the other. These led across the top of the booth and then bent down, so that their ends described a semi-circle, parallel with the central points of the scales. On the end of one arm was a point light source and on the other a sound source. Both arms projected back from their locus a short distance in front of the booth where the experimenter stood. On each of these projections was a pointer which moved around a semi-circular scale divided in the same way as the scale seen by the subject, so that by placing the point to a corresponding portion on the investigator's scale the stimulus could be delivered at any required point on the subject's scale. In the mid-point of each division on the subject's scale a round hole was cut in the paper so that the blanket material showed through. When the light was switched on it appeared as a diffuse spot in the centre of one of these holes.

---

[1] Langfels, H. S., 'Synaesthesia', *Psychol. Bull.*, 1926, 23, 599–602 and 1929, 26, 582–5; Kelly, E. Lowell, 'An attempt to produce artificial chronaesthesia by the technique of the conditional response', *J. Exper. Psychol.*, 1937, 17, 315–41; Howells, S. H., 'The experimental development of colour tone synaesthesia', *J. Exper. Psychol.*, 1944, 99, 660–7; London, I. D., 'Research on sensory interaction in the Soviet Union', *Psychol. Bull.*, 1954, 51, 531–68.

[2] Howells, 1944.

The travelling light source was a directed six-volt lamp. The travelling tone source was a telephone ear-piece being stimulated by a small pure-tone generator located behind the booth. The tone was set at 1,000 C.P.S. and at 50 db. on the tone generator meter—this gave a tone to the subject of quite audible but low amplitude. The experimenter was provided with a 'lectern' bearing the tone and light control switches and a board for recording the subject's responses. Except for the low intensity lamps on the booth the room was in darkness. The room in which the investigation was carried out was sound-proofed and was lined on the inside with sound absorbent board.

Procedure

The subject had previously been given a screening test for auditory defects and had been interviewed concerning any physical disabilities he suffered. Because of the screening procedure involved in admission to the apprentice training school from whence the subjects were taken, very few subjects had to be rejected on these grounds. On introduction into the test room the subject was seated in a small chair facing the inside of the booth and the subject's scale. Care was taken to see that the subject's head was approximately beneath the locus of the scale. The task was then explained to the subject in the following terms.

'I want you to tell me the direction from which a number of tones come. The tone will sound at any of the points about this scale in front of you. You will notice that the scale goes from zero to 75 green and from zero to 75 red. The tone will always come from one of the positions marked, never from anywhere else and the directions will be more or less evenly distributed about the thirty positions on the scale. I want you to tell me each time there is a tone where it comes from by saying the position; for example it might be 25 green, 65 red and so on.'

The subject was then given several practice tones and asked to identify the direction of each one. If he was in error he was corrected for these trials but never again at any other time. When it was clear that the subject understood the instructions, series A was administered. This consisted of twenty-five trials of identifying the direction of the tone. At this stage no light was employed. Series A served the double purpose of giving a base line indication of the subject's ability to make the judgements and to acquaint him thoroughly with the task before the crucial part of the investigation began. During this session the subject was instructed to keep his head still while the tone was sounding and to watch the zero point in front of him. The tone delivered throughout was a double one. Two short pips separated by about half a second were found more convenient than one pip. The first pip served to attract the subject's attention and generally to orient him towards the required direction. Most subjects reported that they based their judgement on the second 'pip'.

When series A was completed the subject was instructed:

'From now on a white light will shine through the holes around the scale. Just as the tone, the light may appear anywhere from zero to 75 green or 75 red. Whenever you see the light I want you to turn your head towards it and to look at it just as you have been looking at the zero. Keep your head still as before, listen for the tone and then tell me where you think the tone has come from. Your task is just the same as before—to tell me where the tone has come from. Since the light may appear at any of the points about the scale you will have to be scanning round to see it. As soon as you do, keep your head still and watch it.' The subject was then given several trials with the light alone so as to enable him to recognize it.

From this point three more series constituted the main part of the procedure. Series B consisted of twenty-five trials in which both the positions of the light and of the tone paired with it, were decided upon separately, by reference to a table of random numbers. Thus each varied randomly about the scale and was independent of the other. The position of the light seen by the subject bore no regular relationship to the direction of the tone which followed it. Series C followed series B without any interruption and the two were continuous as far as the subject was concerned. In Series C, twenty-five trials were given, where the light and tone were delivered from the same position on the scale. Again these positions were determined by reference to a table of random numbers. Dispersed amongst these twenty trials were five in which tone and light were separated as in Series B. The thirty trials of series C thus formed a schedule of partial reinforcement. Series D followed series C also without interruption and again the two were continuous from the subject's point of view. Series D was the same as Series B in that light and tone occurred independently and randomly about the scale.

The whole procedure occupied about half an hour. No time limit was set to the subject's judgements but he was encouraged to make quick decisions by pointing out that his immediate impression of direction was probably the most accurate. As soon as one response was given the pointers of the stimulus arms were set again and a new pair was given. In all cases where the light was employed a period of one second elapsed after the subject had seen the light before the tone was sounded.

The procedure yielded a great deal of data concerning the accuracy of judgements in this type of task and the influence of the light on these judgements when the latter were compared to those in which the light was not employed. However, to restrict the available information to the scope of this section we will deal only with the change in judgements from the series B to the series D.

It was believed that series C would act like a conditioning procedure, that because the light preceded the tone and because the two always occurred together the position of the light would come to signify the direction of the tone. Series B and D were necessary to test the effect of this 'conditioning' series. It was felt that the effect would be indicated by an increased tendency for errors in judging the direction of the tone in series D to be towards the light, wherever the light might be. It was also expected that this tendency would be greater for the group of introverts.

As in the case of salivary conditioning it seemed here a wise precaution to carry out a control study to determine whether the effect of the light on the position of the tone altered with the passage of time simply by virtue of the number of judgements made. Accordingly twenty subjects were put through the schedule with a series randomly independent substituted for series C. Here, no change in the relationship of the light to the tone was made from series B, to series C or from the latter to series D.

Results

As was intimated, this study yielded a good deal of information which will not be included here. Two scores were taken as indicating the 'conditioning' effect of series C. They will be referred to as D–B and No. of +. The first was obtained by calling a positive error, any error in judging the direction of the tone which placed the tone closer to the light than it was in fact and, a negative error, one which displaced the tone away from the light. The algebraic sum of these errors was then computed for series B and D, for each subject and the group mean of sum's difference between the two calculated. This group mean was +79·49 ± 128·82. With eighty cases it yielded a 't' value significant at better than the 1% level ($t = 5{\cdot}5$) suggesting that the sum of the + errors in series D was indeed larger than that of series B.

This seemed to suggest that the procedure was working as was hoped. Nevertheless, it was felt wise to make the same comparison of B and D when no pairing of light with tone intervened. Accordingly the twenty subject control study was carried out. This time, instead of the sum of all the errors in D being positive and those of B being negative as was the case in the main study, both sums were negative, but that of B was the larger quantity, so that the value of D–B was positive for this study also. The group mean difference was here + 147·25 ± 88·13

TABLE 3.7

Pearson product moment correlations between 'D–B' and No. of + and personality parameters

	D–B	No. of +	HY	Pd	E of MPI	C of CPF	Pt	K	'D'	N
40. D–B		+0·531	+0·002	−0·094	−0·163	+0·063	−0·054	−0·085	−0·163	−0·035
41. No. of +			−0·113	−0·125	−0·250*	−0·182	+0·096	−0·103	−0·222*	+0·077

which again yields a '*t*' value, significant at better than the 1% level ($t = 7{\cdot}4$).

The same result was found when the measure taken was the No. of +. This score was obtained simply by counting the number of + errors in each of B and D, i.e. the number of displacements towards the light, and by subtracting the number in B from the number in D. By taking account only of the direction in this way it was hoped to obviate the effect of the occasional very large error. In the main study the group mean difference between these measures was $+1{\cdot}08 \pm 3{\cdot}19$. With eighty cases this mean difference was significant at the 1% level ($t = 3{\cdot}00$), again suggesting that more + errors were made in D than in B.

When the same comparison was made in the control study the group mean difference was $+1{\cdot}4 \pm 3{\cdot}41$. These values yield a '*t*' ratio which is just on the borderline of significance at the 5% level ($t = 1{\cdot}8$).

Table 3.7 shows how these two scores, D–B and No. of +, related to the personality variables.

Table 3.7, in which the principal predictions are tested, shows that only No. of + and E of MPI and No. of + and D yielded significant correlation coefficients.

Discussion

The results of the control study, as in the case of salivary conditioning, renders the meaning of the main study ambiguous. If the results of the control study are to be taken at their face value and, since the subjects were drawn from the same population and the procedure identical, other than the controlled variable, there seems no reason why they should not, then it appears that our procedure measured not a conditioning process but the effect of prolonged exposure to the task; the more often the light and tone were temporally associated the more the light appeared to affect the perception of the direction of the tone irrespective of how the two were related spatially.

The effect seems to us now to be more complex than the original conception of spatial conditioning and would seem to be another instance of the general principle of responding in a situation characterized by scant cues, already noted in connection with verbal conditioning. Other things being equal, this principle appears to assert, events that are associated in time are associated in other ways and the more frequently they are associated in time, again other things being equal, the more closely they are associated in other ways. In our study the effect was, in fact, rather stronger in the main study where the light and tone were paired both temporally and spatially but it was still strong, even in the absence of spatial association.

It is believed that the negative sums of errors were obtained through the overpowering influence of some very large errors which, by virtue of the positions of the tone and light in some trials, were bound to be negative. These errors were associated with trials in both series B and D and at least one instance of such an error occurred with about a third of the subjects. The explanation for these errors might lie in echo effects in the experimental room or might have had something to do with the positioning of the subject. Since no conclusive evidence pointed to these being errors of any different kind to the others it was felt they had to be included. By and large they approximately compensated for each other but, in our terminology they made the sums of errors appear negative except in series D of the main study. It was despite this artifact that the effect of the light on the tone was evident also in the control study, a result suggestive of its strength.

If, in fact, subjects do tend to associate in other ways, stimuli which are only temporally associated when cues are reduced to a minimum, the process could be regarded as reflecting an existing established conditioned response. In the general experience of the developing organism when one thing follows another closely in time, they frequently *are* associated in other ways; one aspect of the sight of a mother, for example, and the sound of her voice, is a purely temporal association. Many other things are coincident in time and are not associated but such sheer coincidences come to be ignored as the result of a discrimination process, one might suppose. In the absence of other cues to effect discrimination, to respond to one stimulus that always precedes another as if they were the same stimulus, is clearly consistent with a conditioning paradigm. The missing element is that of reinforcement. In our investigation there is no ostensible reinforcement. By displacing the tone towards the light the subject gets no reward, avoids no punishment or, as far as one can see, achieves no end at all. If conditioning had occurred by virtue of the spatial pairing alone one would have had an example of conditioning by virtue of contiguity (Guthrie, 1952). If our procedure is simply reflecting a general conditioned response, built-in during development, then one cannot say whether reinforcement has occurred or not. In general one would have expected that reinforcement would have played its part in the establishment of this response.

Turning now to the contents of Table 3.7 one must again report a general lack of sizeable coefficients between these measures and the personality variables. Even on one's revised concept of the process, significant relationships would still be expected. However, only No. of + shows any significant correlations and these are difficult to interpret. One coefficient was predicted, that with the E scale of the MPI but the other is with the lie scale of the CPF and these two personality variables are virtually uncorrelated in this study ($r = -0{\cdot}013$). If the coefficients are reproducible ones then the process must be related to some factor, other than extraversion, related to the D scale. This is not neuroticism as neither E nor D are correlated with N in this study.

The possibilities of this procedure for examining what could be the reflections of a very general conditioned response established during early development seem impressive to the writer, although the experience of this investigation has indicated that improvements in apparatus and technique are called for. Even the nature of the general response, if such there be, is doubtful. It would be quite consistent with such facts as there are, if one were to suggest that the true state of affairs was one essentially of discrimination learning. Such reflections serve to emphasize how far one has to go before the relationship of even the simplest laboratory response to general qualities of behaviour, is understood.

GENERAL CONCLUSIONS AND SUMMARY

In the four investigations reported in this section, the relationship between four associative procedures and measures of the personality dimensions Introversion/Extraversion and Neuroticism has been examined. Of the four procedures, eyeblink conditioning, salivary conditioning, verbal 'conditioning' and spatial 'conditioning', the latter two did not seem to possess the special properties of conditioning procedures, at least in the form in which they were originally conceived. After closer acquaintance with these two procedures it was decided that verbal 'conditioning' (probability learning) could best be conceived as a type of trial and error learning and that spatial 'conditioning' seemed, not to involve the establishment of a new conditioned response but to be eliciting existing conditioned responses established during development. Of the four procedures only eyeblink conditioning and salivary conditioning possessed the properties of conditioning procedures. Despite encouraging results from a pilot study, control studies cast considerable doubt on whether, over the whole group, the performance phenomena under examination in salivary conditioning and spatial 'conditioning' were obtained.

In the case of salivary conditioning a nearly significant difference between the amount of salivation to the conditional stimuli and the 'control' stimuli was obtained in the control study, which difference was in the direction of indicating that the conditional stimuli elicited more salivation in the absence of any reinforcing schedule. Essentially the same result was obtained in the control study of spatial conditioning although this could be interpreted as indicating the operation of the same tendency to respond as was suggested for the main study. We have characterized this tendency as one in which events associated only in time, in the absence of cues to the contrary, are responded to as if they were associated in other ways.

Bearing these complications in mind, the relationship of various 'conditioning' measures to a set of personality questionnaire measures was examined in the form of Pearson-product moment correlation coefficients. The personality scales used were: HY (hysteria scale of MMPI);

Pd (Psychopathic Deviation scale of MMPI); E (Extraversion scale of the Maudsley-Personality Inventory); C (of Cattels' Contact Personality Factor Inventory); Pt (Psyaesthenia scale of the MMPI); K (The Faking good/Faking bad scale of the MMPI); D (the 'lie' scale of the Contact Personality Factor Inventory) and N (the Neuroticism scale of the Maudsley Personality Inventory). The scales of the MMPI had been shown to be related to extraversion and neuroticism in a previous study by Franks, Souief and Maxwell (1959). The intercorrelations of the questionnaire measures in this study did not bear this earlier work out in every detail.

The following significant correlations were found between the performance measures and the personality scales:

Salivary Conditioning

(*a*) $\Sigma \frac{(T - C)}{N}$ (mean difference in salivation between conditioned stimulus and control situation)—with K, $r = -0{\cdot}319$.

(*b*) No. of + (number of times salivation to conditioned stimulus exceeded control measure irrespective of amount)—with Pt, $r = +193$, with K, $r = -0{\cdot}349$, with D, $r = -0{\cdot}222$.

It was concluded, on examining these relationships, that, if any of the indicated results were reproducible, neuroticism and salivary measures were positively correlated. In the absence of any corroborative evidence and in the presence of contradictory facts, however, it seemed most likely that these correlations would not prove reproducible and, in view of the control study results, did not, in all probability, involve a conditioning measure.

Verbal Conditioning

(*a*) Change in left choices (gross change in number of 'left' choices made from first schedule with the 0·5 ratio to the second schedule with a ratio of 7 : 13 lefts to rights)—with Pt, $r = +0{\cdot}241$: with N, $r = +0{\cdot}221$.

(*b*) *Variability* (sum of squared deviations of no. of left choices made over seven blocks of twenty trials from corresponding points on straight line of best fit to these seven points)—with HY, $r = +0{\cdot}218$, with Pd, $r = +0{\cdot}220$.

Because the correlations with Pt and N were such as to suggest a *negative* relationship between change and the scales, it was suggested that a rigidity factor was being measured, rather than one of neuroticism. Because Hy and Pd do not correlate, in this study with extraversion or neuroticism nor with any other performance measure, the meaning of the relationship between variability and these scales remains obscure. On the evidence of Franks, Souief and Maxwell (1959) it suggests that the more extraverted subjects were more variable. This relationship was also obtained in the pilot study.

Spatial Conditioning

(*a*) No of + (count of the difference between the number of error displacements of tone towards a light before and after a series of light and tone being spatially paired)—with E (of MPI), $r = -0{\cdot}250$: with D, $r = -0{\cdot}222$).

Again, since E and D are uncorrelated it is a little difficult to see what their relationship with the spatial conditioning measure could mean. The negative correlation with E was predicted but as it stands alone we must reserve judgement until this correlation is reproduced in another study.

Eyeblink Conditioning

No personality scale in this study correlated significantly with eyeblink conditioning. It has been shown that when normal subjects have been employed this result is consistent with three other similar studies in which no significant correlations are reported. The exception is a study by Franks in which a significant correlation of −0·46 was reported between extraversion (E of MPI) and eyeblink conditioning in a group of 57 students.

A *z* transformation was performed on the coefficients and the estimated true correlation between eyeblink conditioning and extraversion proved to be a low negative correlation significant for the number of cases being considered.

Reasons for low values of the obtained coefficients have been discussed, particularly in regard to unknown environmental conditions and the inadequacy of the principal questionnaire scales, the Maudsley Personality Inventory. The former factors are judged of possible but unknown importance, but evidence is brought to bear on the latter contention which supports the suggestion made that the E scale of the Maudsley Personality Inventory is a poor criterion in differentiating groups of introverts from groups of extraverts.

At the level of analysis attempted in this study, then, our conclusions must be negative. Taken as separate studies the four procedures did not reveal performance scores which significantly correlated with measures of extraversion or of neuroticism in any consistent and meaningful way. The relationships that were found seemed to be of the type that could have been thrown up by chance in a large matrix, at best their meaning was obscure. However, that this study does not conclusively settle the issue of whether conditioning measures correlate with extraversion and neuroticism, is clear when one considers that only eyeblink conditioning and salivary conditioning possessed the properties of conditioning procedures, in the form they were employed and of these, conditioning was only indisputably achieved in the former.

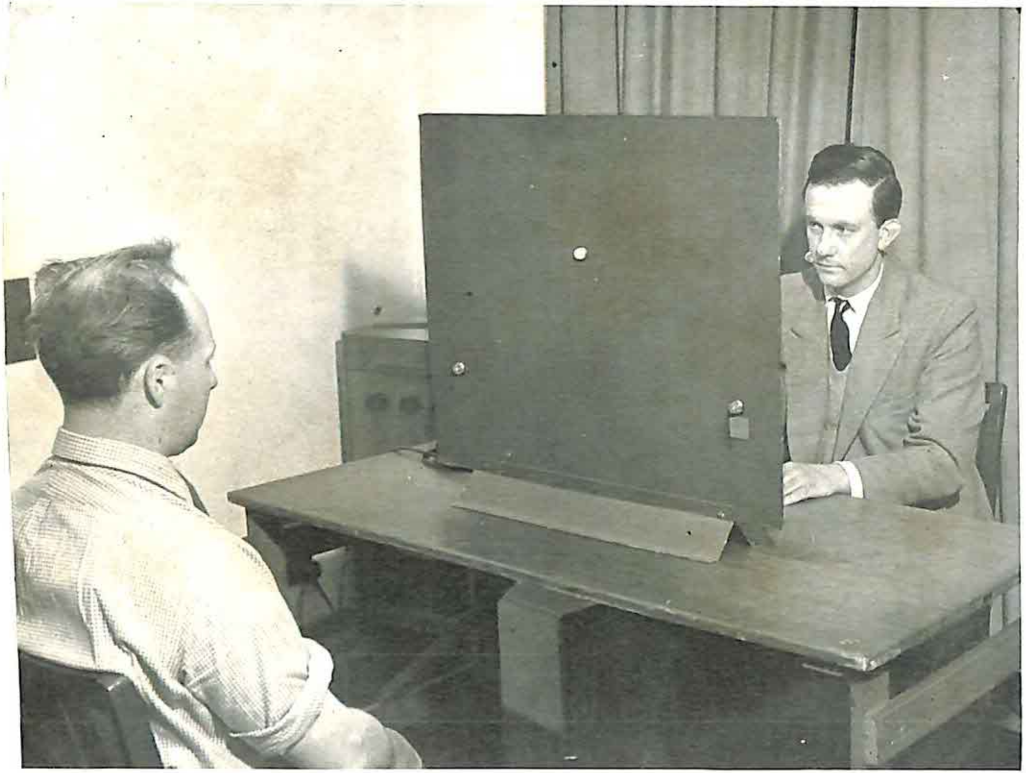


PLATE V. Verbal Conditioning Apparatus. Subject on left, experimenter on right.

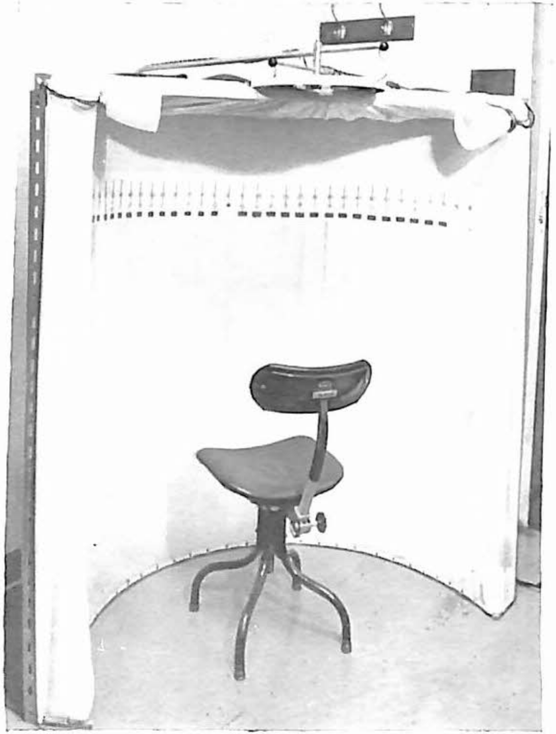


PLATE VI. Spatial Conditioning Apparatus. Showing chair for subject, subject's scale and experimentee's setting scale.

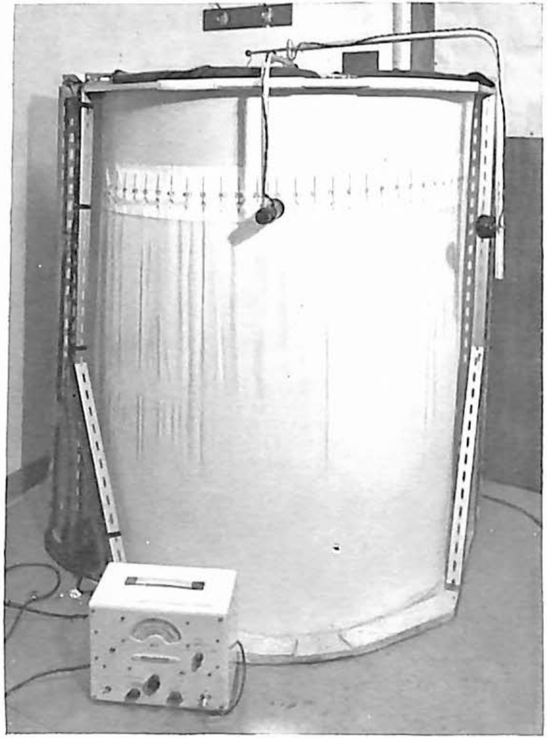


PLATE VII. Spatial Conditioning Apparatus. Rear view showing moving tone and light sources and duplicate of subject's scale for checking the position of moving stimuli sources.

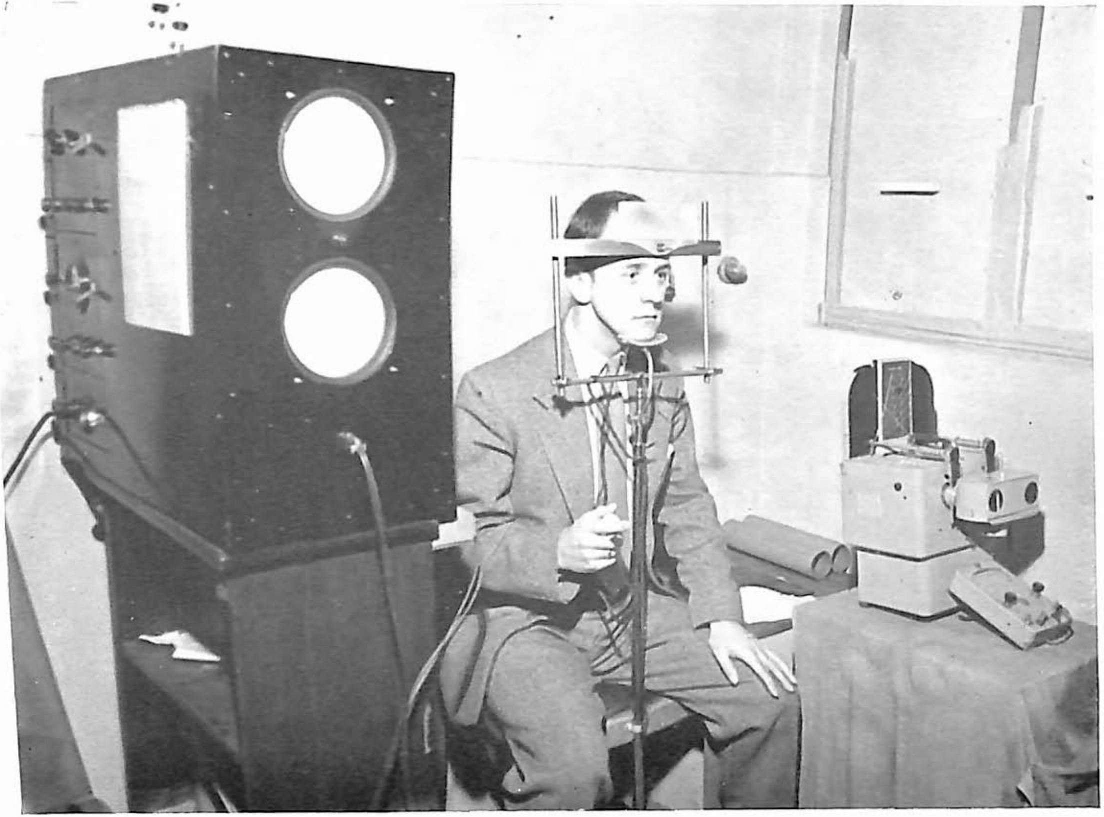


PLATE VIII. Figural After-effect Experiment showing position of subject in the centre, with Comparison of Circles Test on his right and Brightness Comparison Test on his left.

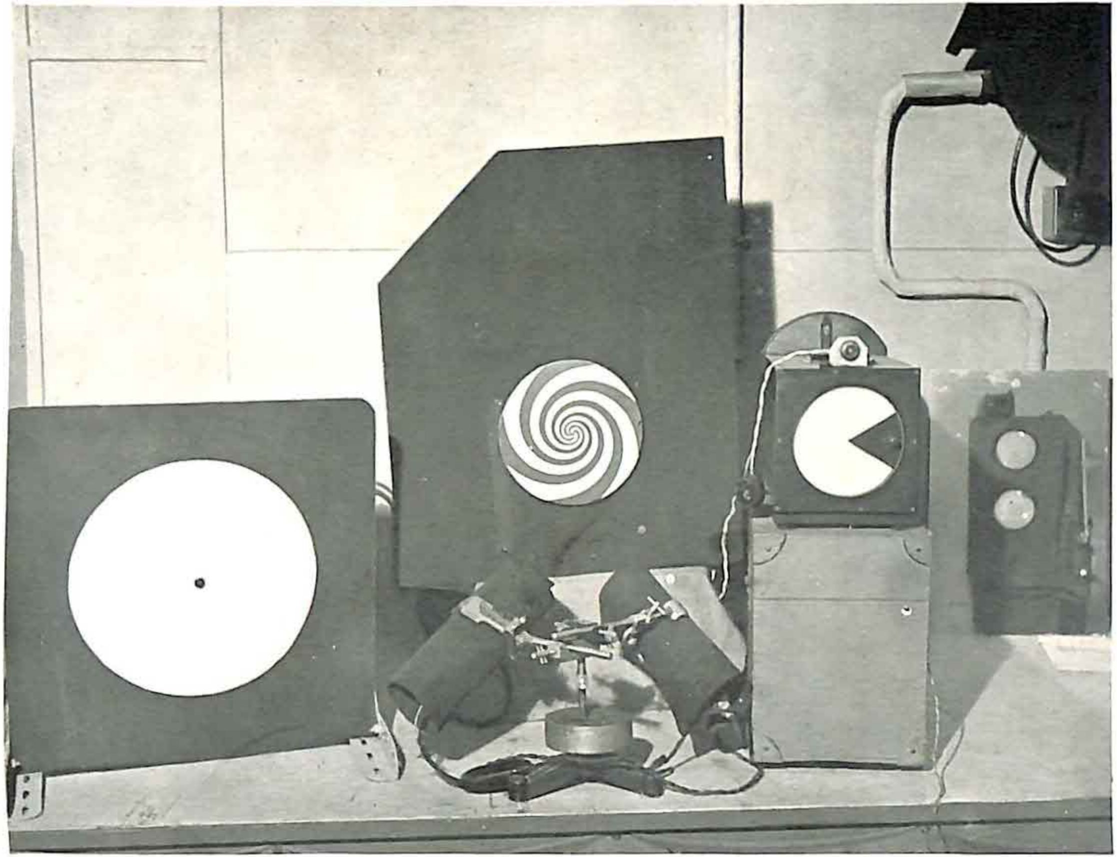


PLATE IX. Subject's view of tests. Extreme right: mirror reflecting Comparison of Circles Test; extreme left: projection screen of Brightness Comparison Test. Centre left: Rotating Spiral Test; centre right: Dark Vision Test.

4

MEASURES OF PERCEPTUAL FUNCTIONS

H. C. Holland

Introduction

SHOULD it prove possible to establish experimentally a general feature of several perceptual tasks which is related to a relatively well defined personality dimension, three things would have been accomplished: (*a*) it would extend the scope of the theory of personality from which the existence of the dimension can be derived; (*b*) it would permit a *rapprochement* between perceptual task performance, and performance in those other sectors of behaviour to which the theory has relevance; (*c*) it would permit 'personality' to be used as an integrative concept in psychology accounting for the differing degrees of generality found in the behaviour of an individual (or type of individual) from one situation to another.

The feasibility of relationships between perception and personality stems from several sources. (I) In the first place, the relationship can be inferred from the establishment of genetic determinants in both perception and personality (McLeod, 1954). Few modern psychologists would deny that such a relationship exists although they may disagree about the contribution made by endowment in either case. (II) A second source of evidence is the large number of investigations which have outlined differences between clinical groups, and perceptual abnormalities of a psychogenic origin. (See for example, Eysenck, Granger and Brengelmann, 1957.) (III) There exist a group of theories aimed at the explanation of changes in performance consequent upon previous stimulation and response, Pavlov (1927), Hull (1943), Eysenck (1953) and Kohler (1944). These are generally referred to as 'inhibition' theories, and are based upon the proposition that a positive force exists which acts against evocation and which is due to either preceding work or excitation. (IV) Finally, there is the direct experimental evidence and general characterizations of perceptual task components (Thurstone, 1943, Eysenck *et al.*, 1957) represented by 'factors' which in themselves reveal nothing of the causal relationship until predicted by a causal hypothesis.

The most important of the sources outlined is the theoretical one. Without theory there can be no causation and only the existence of a theory allows 'causal' statements to be made. The theory within which the facts of the present study will be interpreted is a polygamous association between Pavlov, Hull, Kohler and Eysenck, the theory of the latter having the widest application in the field of personality study and being a development and modification of the theories of Pavlov and Hull. Kohler's theory, on the other hand, has been concerned mainly with changes in perception which follow continued stimulation; although its formal and functional identity with similar concepts in learning and personality have been outlined (Eysenck, 1955) and (Duncan, 1956). Hence differences in the inhibitory substratum which is postulated by Eysenck to account for differences in behaviour are applicable to the perceptual phenomena outlined by Kohler.

According to Eysenck's theory differences in the excitation/inhibition balance of the mechanisms subserving behaviour may be assessed on the behavioural level by introversion/extraversion, and a continuum (or dimension) corresponding to this personality variable is consonant with a continuum of 'inhibition'; the extravert being high on inhibition and low on excitation and the introvert being essentially the opposite. Unlike Eysenck, Kohler has not been primarily concerned with individual differences but has postulated that constant stimulation of a sensory surface leads to a state of localized fatigue (inhibition) in the mediating mechanisms of the cortex which has a measurable effect on subsequent stimuli presented in and near the same area (Kohler, 1944). Nevertheless, Kohler's initial experimental results show a wide range of individual differences and the phenomenon of figural after-effect has been employed quite widely as a measure of inhibition in several pathological and normal groups (*vide infra*).

Now it follows from the postulated identity between other inhibitory effects and satiation and their relation to extraversion, that differences in satiation must be paralleled in some measure by differences in extraversion. Indeed Eysenck, arguing from the same premise, has asserted that satiation effects should appear *earlier*, they should be *stronger* and that they should disappear more *slowly* in a hysteric group of subjects Eysenck, 1955): where the hysteric represents an extreme group for extraversion.

The present study makes essentially the same prediction as Eysenck concerning extraversion in a normal (unhospitalized) group of young males, although several of the perceptual tasks assigned involve changes which are not clearly accountable within the Kohler Wallach hypothesis. The relevance of satiation theory to each test will be discussed in each case.

The Tests

A battery of six tests was employed, each giving a measure which is a function of the existence or build-up of inhibition. Of these six, two were tests which had not been used before, namely a test of visual fixation and a test of brightness satiation with figural after-effect. The remaining tests were dark adaptation, Archimedes spiral, contoured figural after-effect and auditory flutter fusion. The tests will be dealt with separately and the specific predictions outlined in each case.

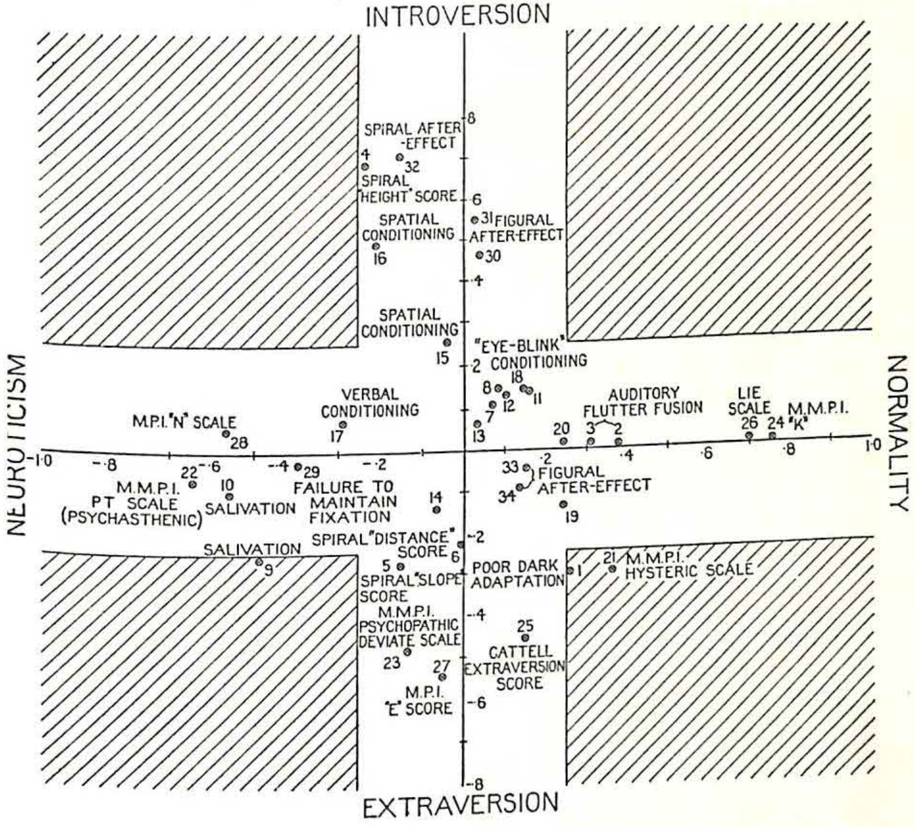


FIG. 4.1.

After each test description are given the two basic examinations of the obtained results. The first examination takes the simple form of the description of the distribution and the comparison of two selected groups assumed to represent high and low estraversion groups because of their scores on the Maudsley Personality Inventory. Secondly, there is an examination of the correlational aspects of the investigation which outlines the relationships between the test scores and the questionnaire criteria as well as the more important correlations between different experimental variables. The factorial analysis of the intercorrelations is dealt with later on. (P. 234; cf. also Fig. 4.1.)

Several derived scores have been used in both of the examinations referred to and it may be useful at this juncture to designate each score and what it represents.

(1)[1] 1. *Dark adaptation.* The score, in minutes, taken by the subject to achieve four correct discriminations at the level noted in the text.

(22) 2. *Auditory flutter fusion.* I. The threshold of frequency discrimination.

(23) 3. *Auditory flutter fusion.* II. The differential threshold for frequency discrimination. Defined by the difference between 0% and 100% 'correct' responses.

(29) 4. *Spiral after-effect* (Height).[2] The maximum after-effect duration. This may be defined as; height; $a = y_1 + y_2 + y_3 + y_4 + y_5$ and is the average value of y calculated as $a/5$.

(30) 5. *Spiral after-effect* (*Slope*). Gives the net increase or decrease in y from the first to the fifth reading. It thus defines the slope of the line connecting y_1, with y_5. Slope therefore $= b$, $b = y_1 - y_5$.

(31) 6. *Spiral after effect* (*Distance*). Defines

$$c = \frac{\{4(y_2 + y_3 + y_4) + 6b - 12y_1\}}{\sqrt{16 + b^2}}$$

and represents the perpendicular distance of the intermediate points, y_2, y_3 and y_4 from the line of the slope: positive when they are above the line and negative below.

(54) 7. *Fixation.* The sum of the two one-minute trials, i.e. the total number of times the subject breaks fixation.

(55) 8. *Figural after-effect* (*Contour*). The difference between the pre and post satiation judgements on ascending trials.

(56) 9. *Figural after-effect* (Contour). The difference between the pre and post satiation judgements on descending trials.

---

[1] The numbers contained in brackets are those by which the tests are denoted in the factor analysis in the section dealing with this aspect of the investigation.

[2] This method of summarizing the spiral after-effect scores was suggested by Dr. Patrick Slater to whom acknowledgement is made for both this and other statistical advice in the production of this report.

(57) 10. *Spiral after-effect*. The sum of two one hundred seconds trials of prior stimulation.

(58) 11. *Figural after-effect* (Brightness). The difference between the first and the third pre, and the fifth and the seventh post satiation judgements.

(59) 12. *Figural after-effects* (Brightness). The difference between second and fourth pre, and the sixth and the eighth post satiation judgements.

(45) 13. *Intelligence* (IQ). The raw scores on group test 90A of the National Institute of Industrial Psychology.

(46) 14. *MMPI*. The raw score on scale Hy.

(47) 15. *MMPI*. The raw score on scale Pt.

(48) 16. *MMPI*. The raw score on scale Pd.

(49) 17. *MMPI*. The raw score on scale K.

(50) 18. *Cattel's (CPF)*. The mean of the 'E' scale scores on both forms of the CPF.

(51) 19. *Cattell's* (CPF). The mean of the 'D' scale scores on both forms of the CPF.

(52) 20. *Maudsley Personality Inventory*. The score on the Extraversion ('E') scale.

(53) 21. *Maudsley Personality Inventory*. The score on the Neuroticism ('N') scale.

Subjects

The total data available are for 97 subjects, divided into two groups of 17 and 80; the smaller constituting the experimental group for the pilot study which preceded the larger. The first sample were all male graduate or undergraduate members of the University of London reading for a diploma in Education. They were all unaware of the purposes of the experiment and were paid for their services. Their ages ranged from twenty to forty (Mean 29·17, S.D. 6·28). The present tests took some one hour and thirty minutes of a seven-hour testing programme. The second group of eighty were all young industrial and student apprentices. They were again all males with ages ranging between sixteen and nineteen. All were employed by the Ford Motor Company who donated their time; the testing programme being carried out on a normal working day. This sample, therefore, received no direct payment for their services, except for meals and expenses.

Differences between the two groups cannot be ruled out. The second group is younger, probably less intelligent, and there may have been a slight difference in motivation. The actual values quoted in the text are for the larger population. The smaller group scores have, where appropriate, been appended in parenthesis.

The test situation

All the tests to be reported were conducted in the same room, which was 9·33 × 10·25 ft. square. It was 8·33 ft. high and thus had a capacity of 796·83 cubic feet. Within the limits of permissible expenditure the room was prepared for the present investigation. Both windows were shuttered and light tight, and walls and ceiling painted matt grey to prevent the reflection of extraneous light from instruments. The edges of the doors were sealed and the doorway was overhung by a heavy curtain. The corridor outside the room was carpeted to prevent excessive noise and a red light warned passers-by that an experiment was in progress. The room was ventilated by an electric fan which maintained a continuous level of sound, and heated by hot water pipes. Both temperature and humidity were kept at a relatively constant level throughout the investigation.

All tests but one (AFF) were so arranged that they could be performed from the same position (see Plates VI and VII). Once the subject was seated in the combined chair and head rest he was not disturbed until the battery was finished. (He could, of course, relax from any position of strain during the rest periods between tests.)

Questionnaires

Several questionnaire measures and an intelligence test score were taken as part of the present investigation.

MPI (*Maudsley Personality Inventory*). This questionnaire has been used extensively as a measure of personality and for a detailed account of its standardization and development the reader is referred to Eysenck (1956), Eysenck (1959) and the review by Jensen (1958). The two scales which have been used are extraversion and neuroticism. The inventory contains 80 questions, 24 being Extraversion, 24 Neuroticism; there is a lie scale of 20 items and 12 buffer questions. Scale items are highly correlated with their factors and insignificantly correlated with others, maximizing the orthogonality (unrelatedness) of the two basic factors which have been given the connotation mentioned above. The lie scale was not used in the present study.

Cattell's CPF (*Contact Personality Factor*). This was the second scale of extraversion. Essentially it is a measure of 'social extraversion' referring to the degree of 'social contact a person needs in his work in order to find it personally satisfying' (Cattell, 1954) and Jensen (1958). There are two forms of the CPF, Forms A and B, which are highly correlated, i.e. 0·86 (Jensen, 1958). In each form there are two scales: E and D. The 'D' scale is essentially a lie scale. Both forms, and the resultant summed scores for E and D were given and taken as the CPF measures for statistical purposes.

MMPI (*Minnesota Multiphasic Personality Inventory*). This third personality test requires no elaboration here. (The interested reader is referred to Hathaway and McKinley (1943) for the uses and outline of the inventory.) Four scales of the test were used, namely, Hy, Pd, Pt, and K, representing neurotic extraversion (Hy), Psychopathy (Pd), neurotic introversion (Pt), and attitude to test items (K).

Unfortunately the MMPI scales, unlike the MPI, are not factorially pure. The work of Comrey (1958) at the University of California has shown that each of the scales used in the present study may be composed of as many as fifteen factors and that many identified factors are common to several scales.

The measure of intelligence available for the industrial group was the NIIP Test 90A. This is a group test of verbal intelligence which is administered to all apprentices of the training school. It comprises four scales, (1) same and opposite; (2) Analogies; (3) Jumbled sentences; (4) sentence completion. There are forty items in the first three scales and seventeen in the last. Raw scores only were used, although the test allows an adjustment for age. As indicated already, there was no intelligence quotient available for the University population.

Results

The means and standard deviations of all the questionnaire scores are given in Table 4.1. These indicate that on the MPI the experimental sample is slightly more extraverted than the general population when

TABLE 4.1

Questionnaire Scores

Means and standard deviations of scales on MMPI, MPI, CPF, and the IQ measure, NIIP Group Test 90A, N = 80

			Mean	S.D.
MMPI	Scale	Hy	16,808	3·824
	,,	Pt	9,667	6·225
	,,	Pd	12,256	3·771
	,,	K	15·039	4·274
CPF (both forms)	,,	E	37·968	4·626
	,,	D	6·448	2·172
MPI	,,	E	29·295	8·361
	,,	N	20·987	10·122
IQ	Raw Score		82·269	13·706

compared with existing normative data (Jensen, 1958), but not more so than normal American University students. The sample is almost at the mean on the neuroticism variable, the mean being derived from a large population from which the sample was drawn. No strict comparison exists for such a homogeneous sample on the CPF and the MMPI although the fact that the first standard deviation on either side of the CPF mean span 'ranks' 5, 6, and 7 support the MPI scores in indicating that the sample is slightly more extraverted than a normal group (Cattell, 1954).

Dark Adaptation

A necessary control which has to be exercised in visual experiments of the kind dealt with in this chapter, is the state of adaptation of the retino/cortical mechanisms. This is particularly the case in figural after-effects and other measures of satiation. Kohler and his associates have argued that satiation is not a transient effect and that a certain amount depends on the permanent state of the (visual) cortex. This residual satiation they call 'permanent' or 'statistical', and it is caused by the nearly constant impingement of visual stimuli upon the receptors. It is a state which will fluctuate moderately as sleep and unusual environmental conditions raise or lower the average intensity of figural currents. Attempts at validating the concept of statistical satiation have not met with any great success, e.g. when differences between early morning and late night measures are taken (cf. also Heron, 1956), but it is consistent with the notion of differential conductivity as postulated by Klein and Kretch (1952).

It is postulated that where resistance is high, figure currents will be weaker, and hence polarization by such current will be less. It would, therefore, follow that the individual in possession of a high level of permanent satiation should produce less effect from the imposition of a standard stimulus, which would represent merely an increment added to the general level, and any displacement of a subsequent stimulus should be less profound.

The influence of statistical satiation can probably be investigated by adequate manipulation of the environment but where this is not possible, it can be controlled, at least to some extent, by trying to ensure that all subjects come to the test in the same state of adaptation. The condition can, of course, only be a rough approximation because of differential rates of adaptation (see Granger (1953), for a good discussion of individual differences in dark vision), but if intensities are set which take only short periods and thus do not capitalize on such differences, the control will be of some value. The time taken to achieve discrimination at a constant intensity may also prove a valuable measure in its own right, particularly in a case such as this where all subjects were seated in a constant room lighting for some 25–30 minutes before dark adaptation was begun, and thus came to the test in rough equality of light adaptation.

A test of dark adaptation, even of the limited type described, in the present battery serves two purposes. In the first place it permits some control of an experimental variable which will effect satiation processes in other tests and secondly, it will act as a measure of extraversion and neuroticism; both dimensions having been shown to have high factor loadings on this function (Himmelweit, 1946). Because of its multifactorial composition the test cannot be expected to correlate very significantly with either of the questionnaire dimensions.

Apparatus

The instrument used in this test is one which was specially constructed for a previous research (Gravely, 1950). It was an Admiralty Research Laboratory Adaptometer Mk. 1a, covering a wide range of illumination from approximately $3{\cdot}961 \times 10^{-7}$ to $7{\cdot}847 \times 10^{-3}$ e.q. f.c. by means of a calibrated light source, three filters, and sixteen graded apertures. The 6 v. D.C. supply was controlled by rheostat and voltmeter. The adaptometer has a uniformly illuminated 5½-in. screen (va. = 4·5°) interrupted by a sector which may be rotated into eight positions. The intensity set in this investigation was $10^{-4} \times 0{\cdot}60$ candles/sq. ft.

A fixation point was attached to the apparatus just above (1½ in.) the illuminated screen and controlled by rheostat to give the minimum position light. The stimulus thus fell in the lower part of the visual field.

The subject was required to fixate the red light during the period of dark adaptation and to report, as soon as he was able, the position of the sector. A successful response led to a change of the sector to another position. Changes were random and four correct responses were required before the test was considered completed.

Results

An examination of dark adaptation scores has shown a wide range from 1·3 to 8·3 minutes. The mean of the distribution is 3·858 minutes with a standard deviation of 1·705 minutes. For the sake of completeness the same two extreme groups which are compared on the other tests were also compared on this measure, although dark adaptation cannot be considered as a simple measure of inhibition. Scores for the two groups were as follows:

Low E: 4·509, S.D. = 1·833, high E: 4·213, S.D. = 1·673. See Fig. 4.2 for the total and selected groups distributions.

These means and standard deviations are essentially the same, the difference between them (0·296) being insignificant.

As was expected, dark adaptation is not correlated with the criterion measures of extraversion. There are, however, several interesting and significant correlations with other test variables, i.e. *height* ($r = -0{\cdot}208$) and both thresholds of the contoured figural after-effect (Asc. r = 0·259) (Desc. $r = 0{\cdot}242$).

The two correlations between dark adaptation and the comparison of circles scores are, of course, consistent with the theory advanced by Kohler and Wallach. That is to say, individuals who take considerable periods to adapt to an initially sub-threshold constant illumination level might be said to be dissipating inhibition which has been generated by preceding stimulation. It can only be assumed that longer time scores represent the relative ease with which satiation potentials are developed or the relative persistence of such potentials once they have been generated. Conditions of both ease of generation and resistance to decay of the effects of stimulation might lend themselves to produce large and persistent figural after-effects.

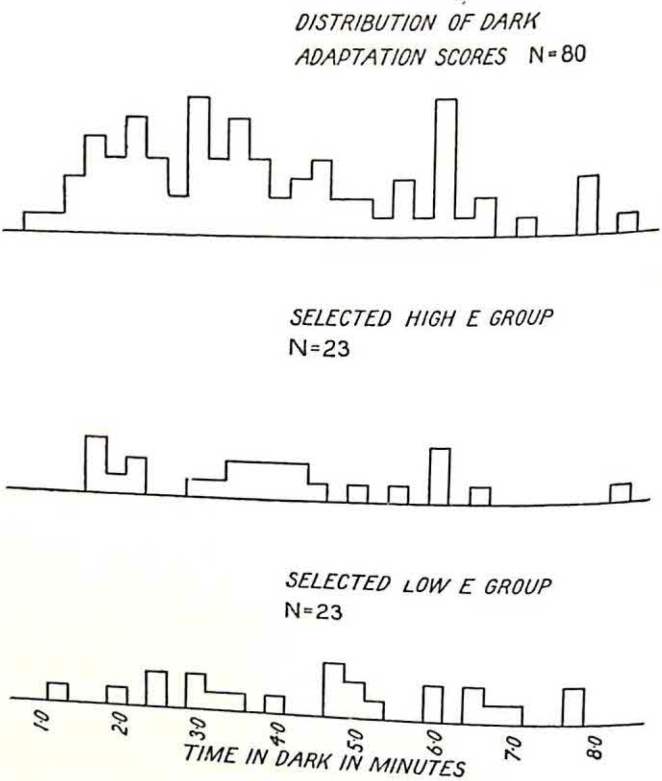


FIG. 4.2.

No obvious explanation can be given for the negative correlations between Dark Adaptation and the derived spiral score *height* ($r = -0{\cdot}208$), and the similar spiral score (100 secs.) ($r = -0{\cdot}107$). The writer's own experiments have indicated the small importance of illumination as a factor in the persistence of the spiral after-effect (Holland, 1958) and the only conclusion that can be drawn is that if the correlations observed are not spurious they must represent some response of the visual mechanisms to the sudden introduction of light which militates against apparent movement after-effects.

Spiral After-effect

The physiological basis of the after-sensation of movement, following the fixation of a rotating spiral or striped drum, is unknown. All that can reasonably be said of it is that it is one of the several illusory effects which are subsumed under the general heading of after-sensations of apparent movement. Several of the factors which influence the occurrence and persistence of the apparent movement have been delineated, and it is known and employed quite extensively because of its simplicity of experimental management and its reliability and scope for research into individual differences with both normal and pathological individuals.

The history of apparent movement certainly antedates psychological literature and probably antedates written history (Wohlgemuth, 1911). In more recent times many well-known figures in psychology have investigated apparent movement (Purkinje, Adams, Muller, Helmholtz, Wundt). These investigators used mainly the 'waterfall' illusion and it is to Plateau (1850) that is accorded the construction of the first arithmetic spiral; although his primary interest lay in the colour effects which can be induced by rotation of the black and white disc.

In the present century the earliest systematic investigation of the spiral phenomenon was conducted by Wohlgemuth in 1911. Outlining the existing theories, which were based on a multitude of different concepts from the 'passage of after-images across the retina' to the 'incorrect interpretation of sensations', Wohlgemuth embarked upon a series of some thirty-four experiments in order to determine the relative effect of contributing variables. Unfortunately, the conclusions which he drew from his investigations have not the clarity of his experiments. Indeed, he postulates a complex network of cortical and subcortical centres, suggests that the spiral after-effect is similar to the spinal reflex, and postulates a 'centre of movement', to account for the phenomenon. Despite theoretical weaknesses the empirical results of Wohlgemuth's work were most valuable in outlining the importance of such variables as brightness, direction, velocity, and area of retina stimulated, and he made the useful suggestion that the after-effect might be employed in the investigation of the 'optic tract'.

Since Wohlgemuth's experiments the spiral phenomenon has been studied extensively by Hunter (1914 and 1915); Thalman (1921); Granit (1928); Grindley (1930); Gates (1934); Deutsch (1950); Walls (1953); Holland (1957, 1958) and Eysenck and Holland (1957), each in turn being interested in a slightly different aspect. These experiments have done little to clarify the nature of the phenomenon, although certain basic parameters are seen to emerge. For instance, little doubt can now be held concerning the central nature of the after-effect (Holland, 1957), (Walls 1953), cf. however, Hunter (1914). Another important parameter is the visual angle subtended. The optimal retinal size appears to be in the region of 2–4° va. 0° pa. (although it appears that this magnitude may be increased without any significant shortening of the effect (Holland, 1958), but it does not seem that this is a rod-cone effect as was previously held (Granit, 1928; Grindley, 1930). Within a wide range the speed of rotation of either spiral or drum appears to have little or any effect so long as the subject is able to discriminate the spiral from its background. Indeed, if Deutsch's 1950 results can be repeated (and the present author has failed twice under slightly different conditions) objective rotation of the spiral disc may not be a necessary condition of the after-effect. Deutsch found that intermittent illumination of the stationary disc not only gave the illusion of rotation but also produced the characteristic after-sensation of movement. (In a private communication Deutsch reports that this effect has also been obtained by Grindley.) The after-effect occurs when the gaze (post-rotation) is projected upon parts of the environment other than the inducing or a similar spiral. Whether the effect occurs independently of a structured stimulus pattern is in some doubt. Holland (1957) has reported that no effect was noted by a small group of subjects when their post-rotation fixation was in a dark grey homogeneous background with low illumination. Grindley and Wilkinson (1953), however, report that an effect was observed when post-rotational fixation was on a white ivorine disc. Fixation seems to be important, despite early studies which showed the contrary (Thalman, 1921). Holland (1957) reports that if the fixation point is changed randomly every five or six seconds, no after-effect follows long periods of stimulation. Finally, the most important source of variation in the after-effect is the time through which the initial stimulation is continued (Thalman, 1921; Holland, 1958). The growth in the persistence of the after-effect is, however, not linear with time and several studies by the writer indicate that persistence scores tend to fall off after stimulation periods in the range 80–100 seconds (Holland, 1958).

Because of the wide range of differences in after-effect persistence the spiral has been employed as a method of reliably assessing individual differences. Perhaps the first systematic investigation in this field is the now famous study by Thurstone (1943). The test (No. 24 in his battery) was unfortunately dropped from his final analysis although it correlated with other perceptual tests like flicker fusion, necker cube, and size constancy measures. The present author has attempted to use spiral persistence as a measure of differences in the excitation-inhibition balance which, on Eysenck's theory (Eysenck, 1955), is the neurological characteristic subserving the differences between introverts and extraverts (Holland and Beech, 1958). Adducing evidence from the relationship between inhibitory effects, their increase with depressant drugs and decrease with stimulants, Eysenck and Holland (1957) have shown that the after-effect is modifiable in accordance with the extraversion-drug hypothesis, i.e. that both extraverts and those 'extraverted' by sodium amytal, have shorter after-effects than introverts or those 'introverted' by excitants.

Several recent studies have attempted to employ the existence or duration of the spiral after-effect as an index of either memory impairment or 'organicity'. This has proved largely unsuccessful in the study of memory (Freeman and Josey, 1949; Standlee, 1953) but, despite methodological weaknesses, somewhat useful in the diagnosis of brain damage, Bender and Teuber (1947, 1948, 1949); Price and Deabler (1955); Gallese (1956); Page *et al.* (1957); Holland and Beech (1958) and Goldberg and Smith (1958). The results of these last named studies have indicated that organic patients perceive the spiral after-effect less frequently and for shorter periods than normals; cf. however, Philbrick (1959) and Johnson *et al.* (1959).

According to Eysenck (1956) and Shapiro (1953), the characteristic differences shown by brain damaged subjects, when compared with normals, are due to an intensification of inhibitory effects; Shapiro preferring the Pavlovian concept of negative induction (the irradiation and concentration of inhibition) and Eysenck a modified Hullian concept of reactive inhibition. Predictively these two notions have many similarities. Indeed, it follows from the theory of personality advanced by the latter author (Eysenck) that brain damaged subjects and neurotic extraverts (hysterics and psychopaths) should have many behavioural similarities (Eysenck, 1955), and thus many of the predictions which may be made of brain damaged patients may also be made of extreme extraverts and *vice versa*. This is a proposition which has been tested and in some measure confirmed by the work of Petrie (1952, 1953 and 1953b). Now it is clear from the established association between brain damage and spiral duration, and between brain damage and extraversion, that the duration of after-effect in a normal population will be proportional to the degree of inhibitory potential with which individuals are endowed and which may be assessed on the behavioural level by their degree of extraversion. At the test level this proposition results in the specific test hypothesis: *The duration of the after-sensation of movement following stimulation by a rotating spiral will be negatively correlated with the degree of extraversion.*

Apparatus (*Spiral*)

The spiral and its driving mechanism has been described fully elsewhere (Holland, 1956) and Holland and Beech (1958). (See Eysenck (1957) for a photograph of the apparatus and disc.) It was a four throw 180° logarithmic type which is sometimes called the Exner Spiral. It was 8¼ in. in diameter and from the viewing distance subtended an angle at the retina of 6·7°. Rotation was supplied by an electric motor controlled through a rheostat. Speed of rotation is maintained at a constant level by reference to a built-in strobe-disc and neon mounted on the drive shaft at the rear of the disc. This method permitted a standard rate of 100 r.p.m. to be maintained within about 1%. Illumination was given by two sources of 60 watts A.C. placed 12 in. below the centre point of the disc. The light from the two sources being contained in tubular housings, which did not allow light to leak backwards in the direction of the subject, and projected it forwards as a slightly eliptoid area a little bigger than the disc. The spiral was held on the rotating shaft by a chromium-headed screw which also served as a fixation point.

Procedure

The subject was seated in a combined chair and chin rest. He had been given, before testing began, a demonstration of the after-effect in the following terms: 'This is a demonstration of a test you will be doing later. I want you to keep your chin on the chin rest and your eyes fixed on this small spot you see shining in the centre of the spiral.' The spiral was then switched 'on' and after about 15 seconds: 'Don't move your eyes . . . in a few seconds I am going to stop the disc from rotating. When I do you will have an after-effect. Don't take your eyes from the centre of the spiral and you will see that the after-effect continues for some time. All I want you to do is to tell me, by saying "stopped", as soon as the effect finishes. Don't take your eyes off the centre of the spiral until I tell you to do so.' At this point the spiral apparatus was switched 'off' and the subject reported the end of the effect. The last injunction of the instructions was added because several subjects in the preliminary studies had been 'fooled' by the very rapid contraction in the early moments following the cessation of the objective rotation and had failed to observe the slower contraction which follows it.

The subject was then asked to describe the phenomenon and the experimenter ascertained that he was in fact perceiving the characteristic after-effect. There was never any difficulty in this respect.

The experiment

In the experiment proper the subject was in darkness adapting to the standard level following the Comparison of Circles Test. During this period the following instructions were administered: 'You will remember the demonstration, given earlier, in which I asked you to report when the after-effect of contraction finished. . . . This is the test proper. It is exactly like the demonstration. Remember to keep your eyes fixed on the centre of the spiral both during and after the objective rotation and tell me as soon as all movement on the surface of the disc completely stops.'

At this point the motor was switched 'on' and when the strobe disc indicated the correct number of r.p.m. the lights in front of the apparatus were switched 'on' simultaneously with a stop-watch. At the end of a hundred seconds the machine was stopped and the subject reported the cessation of the effect. This was the first trial and it was followed by a three-minute rest interval; again in complete darkness. The rest period was followed by a second identical trial.

The second part of the spiral experiment came at almost the end of the battery, immediately following the brightness figural after-effect. The only difference between the trials of the second part and those of the first lay in the fact that stimulation times were not held constant but increased progressively. Thus, the duration of the after-effect was assessed after 15, 30, 50, 80 and 100 seconds of stimulation. The rest periods were also different, and this time only one and a half minutes were allowed instead of three minutes.

Two types of measurements were taken of the after-effect persistence. The first was based upon the two trials of 100 seconds prior rotation. The second five trials gave measures which were taken after 15, 30, 50, 80 and 100 secs. of rotation. This latter method stemmed from the results of previous experiments which had indicated that a plateau of persistence time was achieved as prior rotation periods were increased and that the rate of acquisition to this plateau might represent an individual or personality variable. From these five measures, therefore, there was extracted three derived scores which were labelled *height*, *slope and distance*.

These derived scores are defined as follows. Y1, Y2, Y3, Y4 and Y5 represent the recorded measurements for a subject on the five trials. The five scores were then converted into: (*a*) *height*, $a = \mathrm{Y1} + \mathrm{Y2} + \mathrm{Y3} + \mathrm{Y4} + \mathrm{Y5}/5$, representing the average value of Y; (*b*) *slope*, $b = \mathrm{Y1} - \mathrm{Y5}$ and gives the net increase or decrease in Y from the first reading to the fifth and thus defines the slope of the line connecting Y1 with Y5; (*c*) *distance*, $c = (4(\mathrm{Y2}+\mathrm{Y3}+\mathrm{Y4})+6b-12y_1)/\sqrt{16+b^2}$ and defines the total perpendicular distance of the intermediate points Y2, Y3 and Y4 from the line, it is thus a measure of curvature of the growth function from the slope.

Results

The first of the spiral scores are the two trials to assess the after-sensation persistence following 100 seconds of prior rotation under the conditions previously described. The first trial results in a mean of 24·013 secs. (S.D. = 8·045) and the second trial had a mean of 24·346 secs. (S.D. = 7·836). The inter-trial reliability of the test was $r = 0{\cdot}880$. For correlational purposes the two trial measures were combined to give a mean of 24·179 secs. (S.D. = 7·699). The pilot study sample had a mean of 26·01.

The second set of spiral scores are presented in Table 4·2 which indicates the increase in persistence due to the increase in the time of preceding stimulation. (The pilot study group gave means of: 15 secs.: 15·4; 30 secs.: 19·6; 50 secs.: 19·6; 80 secs.: 21·7; 100 secs.: 21·7.)

TABLE 4.2

Spiral Scores

Table showing the growth of persistence of the after-effect with increasing periods of prior stimulation. Figures in the body of the table are means and standard deviations for the total group: N = 80

Prior stimulation (seconds)	After-effect (seconds)	S.D.
15	15·154	5·871
30	18·679	6·210
50	19·846	5·716
80	20·000	5·800
100	20·744	6·658

The final measures are the derived scores of *height, slope* and *distance*, seen in Table 4.3.

TABLE 4.3

Spiral Scores

Table of derived scores height, slope, and distance. Figures in the body of the table are the means and standard deviations for the total group: N = 80

	Mean	S.D.
Height	18·886	5·370
Slope	−5·590	5·253
Distance	2·817	5·745

The scores indicate again that there is a gradual increase in persistence with increased time and that the small positive score representing the curvature of the acquisition time shows that the curve is slightly negatively accelerated.

TABLE 4.4

Spiral Scores

Means and standard deviations of the second five scores for two selected groups representing high and low extraversion

	1	2	3	4	5
High E	13·1	16·5	17·5	17·5	17·5
S.D.	5·5	6·5	5·0	5·8	5·8
Low E	15·9	20·2	19·6	20·1	20·8
S.D.	6·4	6·8	6·4	5·7	6·2

N = 23 for both groups = 28·75% of total group.

The extraversion hypothesis has been examined by selecting two groups which have high and low scores on the criterion measure of extraversion. Both groups number 23 and thus represent the upper and lower 29% of the total sample; the high E group having scores above 36 and the low E group having scores below 25 on the extraversion scale of the MPI. The scores of the two groups are given in Table 4.4 and the difference between the two groups ($t = 1{\cdot}16\ p < 0{\cdot}05$ n.s.) is also illustrated in Fig. 4.3. See also Fig. 4.4. for the distributions of the total and selected groups.

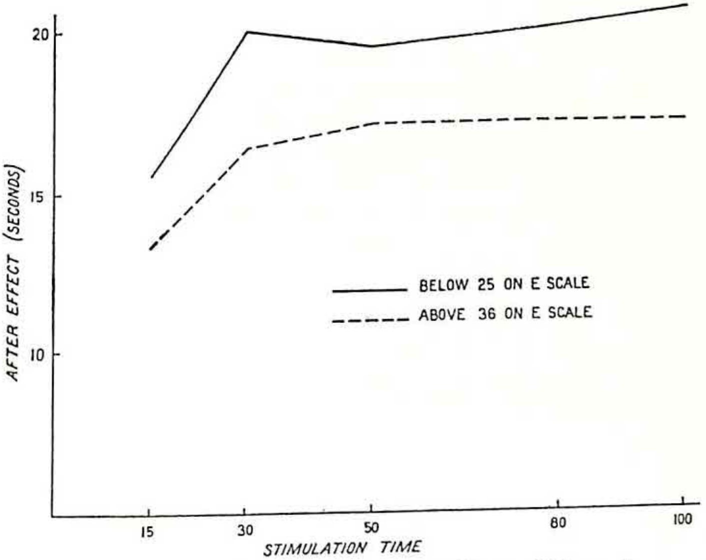


FIG. 4.3. Spiral After-Effect Persistence (Selected Groups).

The results of the spiral test have tended to confirm previous experimental results and generally supported the proposition that the persistence of the after-effect is related to the personality dimension of extraversion/introversion as measured by the MMPI. When extreme groups in a normal population are opposed, discrimination is fairly clear, although not significant statistically, and there is no overlap in the means. It could be expected that pathological groups, e.g. hysterics and brain damaged, would be differentiated even more clearly.

Although test/retest can only be assessed on a trial to trial basis it is very high for this kind of perceptual test, i.e. approximately 80% of the variance between trials is accounted for. The derived score *height* also correlates with the first spiral measure to the degree of $r = 0{\cdot}801$, which, because *height* is an average score of the acquisition rate suggests that this rate is fairly monotonic. This result, of course, also implies that the experiment has not been completely successful in demonstrating the expected plateau of persistence times which had characterized earlier work. Fig. 4.3 suggests that a plateau has already started with the high **E** group but that the low E's may require longer periods still.

In outlining the correlational aspects of this and subsequent sections of the present report, it is not the intention to advance 'strings' of coefficients of every test with every other test. The total matrix of correlations is presented later on and includes the present tests as well as others. All that it is useful to display are the correlations of the specific measures with the independent criteria of extraversion, i.e. MPI and CPF E scores, and any correlations with other test measures which may be of special interest.

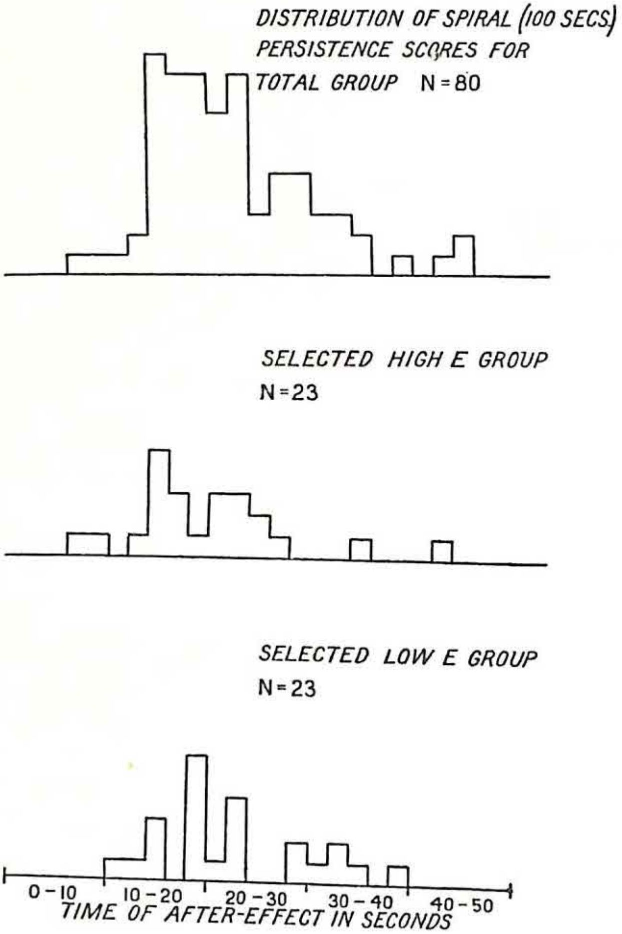


FIG. 4.4.

Four measures of the spiral scores have been used for correlational purposes. These are *height*, *slope*, *distance*, and spiral effect during the 100-sec. trial. The first and the last of these scores are highly correlated, i.e. 0·801, and appear to be measuring the same function. They can, therefore, be considered together and be representative of the persistence of the after-effect in time. The correlation of this function with both criterion scores of extraversion is as follows:

	MPI (E)	CPF (E)
(1) *height*	−0·203*	−0·205*
(2) Spiral 100	−0·154	−0·206*

all are as predicted and three are significant at the 5% level of confidence. (For 78 d.f. 5% = 0·187, 1% = 0·264.)

The correlations of the other scores are not as clear as those in the preceding paragraph, i.e. CPF and *slope*, r = − 0·143; MPI and *slope*, r = 0·069; CPF and *distance*, r = 0·037; MPI and *distance*, = 0·083. None of them achieve statistical significance and three of them are so meagre as to preclude any discussion of them in terms of the hypotheses. The fourth CPF and *slope* suggests that the rate of acquisition of the spiral effect is not related to extraversion, although this conclusion can only be tentative.

Of particular interest are the correlations of *height* and the second spiral persistence score (i.e. spiral 100) with other questionnaire scales, specifically with the Pd scale of the MMPI r = −0·210 and, −0·263 and, in the former case, the K scale of the same questionnaire, *height* and Kr = −0·207. The failure of the measures to display a significant relationship with the Hy scale (neurotic extraversion) Hy and *height*, r = −0·095; Hy and spiral 100 r = −0·147, cannot be accounted for, unless the reason lies in the multifactorial composition of the scale as already outlined.

Of particular interest also is the set of correlations between the two spiral persistence scores and the contoured figural after-effect. They are positively correlated and three of the four coefficients are significant, i.e. *height* and figural after-effect = 0·213 and 0·261; spiral 100 and figural after-effect = 0·163 and 0·252. This suggests, as is pointed out in a subsequent section, that the magnitude of the figural after-effect by the present method of elicitation, is not a simple measure of extraversion, and, like the spiral, is a negative correlate of 'E'.

Fixation

The importance of fixation as a variable in experiments on visual figural after-effects has been emphasized by many studies (Nichols, 1956; Eysenck, 1957; Summerfield, 1955; and Heron, 1956). Yet it is strange that no investigator has reported on the various measures of after-effect in relation to accuracy of fixation, using accuracy in its widest sense to imply maintenance as well as excellence. Too often in psychological experiments, either implicitly or explicitly, the 'co-operativeness' of the subject is stressed and from this the naïve assumption made that because the subjects are instructed to keep their eyes on a certain spot (or fixation point) the effectiveness of the stimuli they receive is equal (cf. Barlow 1952; Ditchburn and Ginsberg, 1952; 1953; Westheimer and Conover, 1954). The vast body of knowledge concerned with individual differences in perception suggests that such an assumption cannot be made and that the relative ability to fixate may be a contributing factor to such differences (Holland and Beech, 1958; Ratliff and Riggs, 1950). (It is true that in a random sample of the population, fixation, as a variable, might be normally distributed, but where different personality groups are chosen or pathological conditions opposed, it may assume considerable importance.)

As Osgood (1953) has pointed out, the fundamental fallacy concerning fixation is the notion 'I fixate (look at) something intentionally by an act of will', for although gross hunting movements of the eyes may be voluntary, the precise alignments which guarantee single vision are due to involuntary mechanisms. Hence volition can play only a small part.

The simple neurological model for fixation by which distributions of excitation representing the disparate images overlap until, reaching a critical point, they summate with lateral inhibition erasing secondary contours, is confounded by perceptual and attitudinal factors, e.g. familiar figures 'fuse' more readily, as do figures having 'good' form, after fusion these figures also resist diplopia. Blinking[1] also involves a break in fixation for not only is vision occluded by the eyelid but, due to Bell's phenomenon, the eyes are moved (Bergman *et al.*, 1952; Bruell and Albee, 1955; Ponder and Kennedy, 1927; Hoffman, 1946). Both eye movements and blinking have been studied after the induction of depressant drug (Bergman *et al.*, 1952; Nathanson *et al.*, 1953), which is of special interest here because of the extraversion/drug hypothesis. Under the influence of the drugs both convergence and divergence of the eyes is altered and in some of the cases reported subjects were able to fixate with one eye only. Indeed, in some of the extreme cases the drug produced both vertical and lateral nystagmus, and other signs of cortical and subcortical sites of action. Any test result which might indicate when the stimulus, in relation to its retinal and central location, was moved, or when it was occluded by blinking, would be a useful addition of information in accounting for individual differences in perceptual phenomena of an inhibitory type.

The Spiral as a Test of Fixation

A recent observation by the author gave grounds for the belief that the spiral might be used as a simple test of eye movements. Under normal conditions, when the spiral is being used as a method of inducing apparent movement, it is rotated at speeds varying between 50 and 150 r.p.m. These speeds are necessary because the after-effect tends to break down beyond them. It has been observed, however, that when the spiral was rotated within the region of 600 r.p.m. and viewed from distances between 3½ to 5 ft., then eye movements whether involuntary or voluntary, or whether due to part of the blink process, produce on the uniform greyness of the disc a 'flash' in which part of the spiral emerged and could be seen clearly for a fraction of a second. The impression gained is as if a part of the spiral were seen as standing still or, in some cases, as travelling at a very much slower rate. The effect also occurs when deliberate blinks are made.

---

[1] Blinking has also been employed as a measure of visual efficiency. The interested reader is referred to the following reports: Luckiesh (1937, 1946); Bitterman (1936, 1946) and Wood *et al.* (1950).

The frequency of the reported 'flashes' varies from about 3–50 (the vast majority being between 5–35), over a one-minute period, but any one person is reasonably reliable from trial to trial.

The following explanation of this phenomenon is tentatively suggested; although it is unlikely that at this stage of development any statement could be comprehensive. If the spiral is rotated at a relatively high speed, but one which will permit discrimination between spiral and background, it quickly loses the appearance of a spiral turning outward or inward on itself and appears as almost concentric 'pulses' moving from centre to periphery. Now, if it is assumed that the same process continues, as the speed increases, beyond the discrimination level, then any change in the fixation point (located at the centre) may momentarily synchronize with the 'expanding' line of the spiral. This stroboscopic effect is only likely to occur when eye movements are relatively rapid and not when they are slow; which may account for the fact that aimless and slow movements do not necessarily produce the phenomenon.

It is rather difficult to describe the characteristics of the 'flash' but many subjects have said that something like 'a black line' appears, or that it 'becomes more black and white' or 'it seems to stand still for a moment'. It seems as if, for a very small moment of time, a certain part of the spiral is suspended in its movement. Not all of the 'flashes' are in the same magnitude or in the same place which suggests a differential direction and amplitude of eye movements. A simple and analogous phenomenon is found when a tuning fork is set to vibrate at a frequency above the discrimination threshold and is moved across the field of vision. The movements of the eye following the fork, being largely ballistic in nature, make rapid jumps which overtake the fork. During the co-incident movements of eyes and fork the prongs are occasionally seen as stationary. It will be appreciated that with the expanding spiral such co-incident movement is made necessary by the fact that *any* movement away from the central fixation point, if its speed and amplitude are great enough, must 'overtake', or be itself 'overtaken' by a part of the spiral as it unwinds.

Preliminary experiments using the spiral as a test of fixation have been carried out. So far they have shown large individual differences with good reliability. In one experiment high correlations (rank order) have been determined between performances on the spiral and reversal figures on the Necker Cube Test, although the relationship is not one to one which militates against a simple hypothesis of reversal being 'triggered off' by eye movements (Eysenck *et al*, 1956). Another set of experimental results from a preliminary study had shown a clear correlation between persistence of after-effect duration and eye movements as measured by the spiral test, i.e. low persistence times went with high number of 'flashes'.

Any explanation of the relationship between poor fixation and low persistence scores is complicated by the yet unexplained nature of the spiral phenomenon and until this doubt is resolved and the effect subsumed under some theoretical generalization, only the empirical data are relevant. However, if the after sensation of movement is properly accountable within a satiation-type hypothesis, the importance of fixation is clear.[1] (The writer has proposed that a possible explanation of the poorer performance of brain damaged subjects on the spiral test (see previous section) may be due to the known distractability of this group. As distraction in visual perception is likely to involve fixation (as well as attentional) efficiency, the possible contribution of its influence must be considered.)

Procedure

To perform this test the subject was seated in his combined chair and chin rest. In his right hand he held a very lightly spring-loaded push button through which circuit to a counter was made or broken. The spiral dimensions, etc., were as for the spiral test. A demonstration was given until the subject was well aware of the phenomenon he was to report and afterwards a practice trial was undertaken. Both demonstration and practice occurred immediately after the demonstration and practice of the spiral after-effect. The subject was then asked to keep the phenomenon in mind until required.

In the test proper, the method called for two one-minute trials with a three-minute interval.

Results

The mean number of eye movements assessed by the present method appears to be very stable over trials separated by only a few minutes. This is shown by the trial means of 20·321; S.D. = 10·158, and 21·218, S.D. = 10·182 and the inter-trial correlation of 0·937 (Test mean for pilot group = 18·8). On the grounds of these scores there can be no

[1] Indeed, in one experiment it was determined that if the fixation point were changed every five seconds, while attempting to induce an after-effect on the spiral, no effect, or a very much reduced effect, followed considerable periods of stimulation.

doubt concerning the *reliability* of this test of fixation although its validity cannot be adequately assessed. Content validity would appear to be high because of the logic of the test situation and because it is hard to imagine what the test is measuring so reliably if it is not eye movements. The predictive validity of the test must be considered low because of its failure to correlate highly with other tests which might be affected by eye movements. Furthermore, when the extreme groups which have been alluded to previously are compared, the measure fails to discriminate them at an acceptable level of confidence; i.e. Low E, 18·717, S.D. = 8·57; high E, 21·087, S.D. = 11·71, a difference of 2·370 ($t = 0{\cdot}766$, p n.s.).

The examination of the correlation of this test with the criterion measures and other visual tasks indicates its failure to behave in the predicted direction, and support other experimental indications that it is a correlate of extraversion and has a considerable part to play in visual tasks requiring steady and maintained fixation.

The two correlations of this variable with the contoured figural after-effect scores, i.e. −0·006 and −0·101, and those between fixation and the main spiral scores of *height* and spiral 100 of 0·004 and 0·012 preclude the possibility of any conclusion. The only other correlations of interest are between fixation scores and some of the questionnaire measures. For example, the Distortion 'D' scale of the CPF ($r = -0{\cdot}238$), the Neuroticism scale of the MPI ($r = 0{\cdot}168$), the K, Pd, Pt and Hy scales of the MMPI $r = -0{\cdot}110$, $-0{\cdot}067$, $0{\cdot}178$ and $-0{\cdot}133$. As several of these scales have a neuroticism component, it is possible that differential eye movements are due in some way to the neuroticism factor.

Figural after-effects

Experience and experiment teach that sensory intensities vary under conditions of continued stimulation. At the simplest level, this is measured as differences between, or progressive changes in, absolute and difference thresholds for stimulation. That a process similar to this sensory adaptation is operative in more complex experience is suggested by a group of findings culled from many diverse areas of psychology. Called severally, adaptation with negative after-effect, figural after-effect, etc., these discoveries indicate that adaptation is a function of stimulus intensity and differential adaptation in a stimulus array due to differential excitation by preceding intensities. Now, as the effects of stimulation persist after the removal of the eliciting stimulus, it follows that a constant stimulus which impinges upon, or crosses, an area which has been subjected to greater excitation than an adjacent area will have less effect in sensation, than in an area of lessened adaptation. In other words, sensory thresholds being a function of the degree of adaptation will be higher in areas of previously strong stimulation.

A great number of early psychophysical experiments have outlined the characteristics of sensory adaptation of the simpler kind and generalizations about its temporal course and other features have been laid down. In the more complex situations, observed in the distortion of contours and the subjective normalization of objectively distorted percepts, two main theories have been advanced to explain the phenomenon. The first (Gibson, 1937 and 1937b, 1954;[1] Gibson and Waddell, 1952) postulate sensory adaptation with negative after-effect as the initial cause of distortion and argue that normalization is due to the possession of the perceiver of certain psychological norms or acquired standards of, say, verticality. If, according to this view, a stimulus is presented which deviates from the norm, adaptation operates to make it like the norm, and subsequent stimuli in the same area are distorted as the complementary of this adaptation. The second theory, and probably the more important, has been advanced to explain a great many perceptual effects in several modalities. Originally suggested by Kohler and Wallach (1944) to account for phenomenal displacements in visual perception, it required a sequence of two stimuli known as the inspection or satiating figure and the test figure. When presented in the correct temporal and spatial relationship, the contours of the test figure are distorted in a predictable way by the inspection figure. This distortion or displacement is called a figural after-effect.

The underlying mechanism of figural after-effects is *satiation*, a condition of the mediating neurones akin to a localized fatigue effect in the central projections of sensory surfaces. One of the basic assumptions of the theory is that a stimulus figure is iso-morphically projected on the cortex setting up, coincident with its contours, a state or condition of increased resistance to subsequent stimuli in the same area. This *anelectrotonus* lowers the *polarizability* of the affected membrane and any subsequent figure current in the same area is deflected away from the satiated region and into an area of less resistance or higher response potential. The resulting perceptual changes are referred to as *figural after-effects*. Satiation as such has the primary logical status of an intervening variable in that no orthodox neurophysiology will cater for it adequately. The original work of Kohler and Wallach contains an extensive discussion of the electrophysiological and chemical substratum which are postulated as subserving the distortions which have been referred to.

Although of primary interest in visual perception, where many parameters of satiation have been established, e.g. temporal aspects by Hammer (1949), Ikeda 1953a, 1953b), Oyama (1953),[2] Parducci (1956), Sagara 1957); Universality by Marks (1949), George (1953a, 1953b);

---

[1] Gibson's experiments are given primacy here but it is clear that he was aware of the work of Stratton on adaptation to reversed fields which must be considered as the forerunner of normalization and similar experiments.

[2] For recent, and some of the most intensive, work on figural after-effects in Japan the reader is referred to the review by Sagara and Oyama (1957).

Retinal size by Sutherland (1954), Prentice (1947, 1950) and brightness by Kohler (1951) and Freeburne and Hamilton (1949), the theory has been successfully applied to other modalities and across modalities (Jaffe, 1956).

In auditory phenomena an effect analogous to the classical visual effect has been described by several authors (Deutsch, 1951; Jones and Bunting, 1949; Krauskopf, 1954; Christman, 1954; Nichols, 1956) and kinesthesis by (Jaffe, 1954; Klein and Kretch, 1952; Eysenck, 1955; Nichols, 1956 and Nachmias, 1953). In autokinetic movement Crutchfield and Edward (1949) have demonstrated the effects of pre-satiation, Carlson (1953) and Hochberg (1950) have shown that prior satiation influences the direction of reversible figures, Deatherage and Bitterman (1952), Shapiro (1954) and Brenner (1953), have outlined its influence in apparent movement, and Kohler and Fishback (1950a, 1950b) and Kohler and Wallach (1944) have maintained that the disappearance of the Muller-Lyer illusion upon repeated trials may be accounted for by differential satiation (but cf. Eysenck, 1958). The crossmodal study referred to above is that of Jaffe who has shown that visual fixation of different widths of paper exerts an influence upon kinesthetic judgements. This finding of Jaffe's is in line with work being conducted in Russia according to Professor Luria of Moscow.[1]

It has been repeatedly observed that there are individual differences in the components of figural after-effects, e.g. magnitude, ease of elicitation, and resistance to decay. Few studies, however, have deliberately investigated these differences. Exceptions are the work of Klein and Kretch (1952), in kinesthesis who have modified the original satiation theory to a proposition of the basal level and state of cortical conductivity and have postulated, and demonstrated, differences between brain damaged and normals. Almost simultaneously with the Klein and Kretch investigation a study by Jaffe (1954), which had several weaknesses, failed to discriminate between a brain damaged and normal sample.

Using the method and apparatus of Klein and Kretch's experiment, Eysenck (1955) made the same prediction of the hysteric as the previous authors had made of the organics. These predictions followed directly from his theory of personality and empirical evidence of the many similarities in the behaviour of these types. The results of the study supported the conclusion that kinesthetic figural after-effects in the hysteric group appear more quickly, are more strongly marked, and disappear more slowly than in the dysthymics. Eysenck's findings on this were supported by Nichols although his very large battery of satiational indicants which included tests of the Muller-Lyer illusion, reversal of perspective, negative after-image duration, apparent movement, and figural after-effects in auditory and visual modalities, failed on the whole to differentiate between his groups of extraverted and introverted neurotics. Nichols' second prediction, i.e. that there would be a significantly high positive correlation between the tests mentioned within each group also failed.

---

[1] This information came from a discussion with Professor Luria on a recent visit to London.

Individual differences in figural after-effects have also been investigated by Wertheimer (1954a, 1954b, 1955a and 1955b). In a series of experiments he has adopted a somewhat different line from previous authors in this field. Beginning from the premise that differences in figural after-effect must represent differences in the modifiability of cortical conductivity he argues that such differences must be related to the general metabolic efficiency of the individual. He has not only been able to establish this assertion (*a*) by demonstrating the relationship of FAE and metabolic drugs, but also (*b*) that schizophrenics, who have low metabolic efficiency, also have small FAE.

Such a comprehensive theory of brain action in perception has, of course, not gone unscathed nor uncriticized. Smith (1948) has posed several questions which satiation theory cannot as yet answer. Hebb (1949) has criticized the anatomical model necessitated by the phenomenon, while, on the neurophysiological level, Lashley *et al.* (1951) has conducted experiments to show that cerebral integration is not due to molar patterns of electrical stimulation. Finally, Luchins and Luchins (1953a and 1953b) have remarked on the inadequacy of the Kohler Wallach theory to account for test figure, and the pattern of which it forms part, i.e. their subjects obeyed the principle of *Pragnanz* instead of demonstrating the Kohler effect.

An alternative to satiation has been presented in the form of the Statistical Theory (Osgood and Heyer, 1952; Marshall and Talbot, 1940). Based on the notion of excitatory 'peaking' in the higher centres due to the different cellular ratio between retina and visual cortex and physiological nystagmus, statistical theory is able to account most of the visual phenomena reported, and possesses the advantage of an orthodox neurophysiology. Although adequate to explain visual figural after-effects it is difficult to envisage the application of statistical theory in other sensory modalities, or even to some visual phenomena (Eysenck and Holland, 1958), see also Smith (1952).

In the present experiment the predictions follow from the same rationale as already outlined for the Archimedes spiral effect, i.e. through the formal and functional identity of satiation processes and reactive inhibition, extraverts should show larger and more intense figural after-effects than introverts, and in a normal population these measures should be proportional to the degree of extraversion assessed by questionnaire.

It is, of course, necessary to qualify the prediction in the light of what is known about the importance of the equality of the stimulation received by the subjects. Eysenck (1957) has already asserted that differential eye movement and the ability to fixate, attention in its widest sense, and the time through which stimulation is continued, will interact to make the figural after-effect a complicated measure.

In this investigation an attempt has been made to obtain a measure of eye movements and the time factor has been controlled by using a constant period. There is, of course, no way of knowing in advance an individual's response to such a time interval.

The preceding prediction was examined in the present investigation on two measures of figural after-effect. These, for the purposes of the present experiment have been called the 'contoured figural after-effect' or Comparison of circles test, and the 'Brightness figural after-effect' or Douple D. From both may be obtained scores which are measures of the magnitude of the influence of the satiating stimulus upon the subsequent test stimulus. This is indicated in the first case by the apparent contraction of the circle which falls within the contour of a satiating figure which is similarly a circle in the same plane and retinal location of a slightly larger diameter. In the second case the score is a brightness matching error due to the continuous inspection of a preceding area of illumination.

The nature of the tests are clear from the descriptions of the apparatus and those of the respective procedures. For the purposes of clarity these descriptions have been presented separately for the two tests. The results are given jointly because they are predicted by, and thus bear upon, the same hypothesis.

Apparatus

The comparison of circles test apparatus comprised two VCR 97 cathode tubes upon each of which had been imposed a circular time base. Both tubes had independent size, focusing, and brightness controls. Fixation was provided by a leim ruby glass filter over a 75–80 micro amp. neon lamp situated midway between the circular tubes. When accurately centred at the correct size (6 cm.) the complete visual array subtended an angle of 3·6°. This visual angle was obtained by having the apparatus reflected in a front surfaced mirror placed in a position 8′ 4″ from the subject and 6′ 10″ from the apparatus. The upper circle was standard figure and the lower the variable. Apart from the size and brightness controls already noted, controls existed for the lateral positioning of the circles in relation to each other. Discrepancies between inspection and test figures were represented by a voltage difference and read off from a voltmeter which permitted accuracy to the nearest millimetre.

Procedure for comparison of circles test

The subject was seated at the side of the apparatus close to the remote control for the variable figure. He was told to keep his chin on the chin rest and to observe the apparatus which was reflected in the mirror directly in front of him. He was then instructed as follows: 'I want you to listen carefully to what I am going to say. In front of you you will see a mirror in which is reflected the piece of apparatus we will be dealing with next. On it you will see two circles, the lower much smaller than the upper. At your right hand you will find a small hand control which changes the size of the lower circle. For the time being I want you to practise making the two circles equal in size, telling me as soon as you have done so.' At this point the subject changed the size of the lower circle, practising to get the 'feel' of the control. 'Now then, in that practice period you very probably tended to look up and down at the two circles to see if they were matched equal. However, in the experiment which follows in a moment I want you to keep your eyes quite still and fixed upon this small red spot (pointing). Do not move your eyes when making the two circles equal. You will find that the distance from this piece of apparatus to the mirror, and from the mirror to your eyes has been carefully calculated to enable you to see both circles when you look at the red spot without moving your eyes.'

The apparatus was switched 'off' at this point, and the subject was given another dark adaptation test. As soon as this test was over, the first part of the experiment was started. The two circles were switched 'on' and instructions given to make four matches of equality, two ascending and two descending.

The four judgements obtained in this way constituted the four pre-satiation PSE's. The post-satiation PSE's were obtained as follows: 'The second part of this experiment is also quite simple, I shall ask you to keep your eye fixed on the red spot in the centre of the apparatus for the next four minutes. The top circle will be "on" . . . but I want you to ignore it and keep your eyes on the red spot. At the end of the four minutes, which I shall call out as they pass, I shall switch "off" the top circle and immediately switch both top and bottom circles back "on" again. When I do that I want you to make four judgements of equality in exactly the same way that you did earlier. Is that clear?' The upper circle was adjusted to give a size of seven centimetres, the subject alerted, and the apparatus switched 'on'. The time was called out as it passed as 'so many minutes gone so many to go . . . please keep your eye on the spot'. Thirty seconds before the end of the total trial period the subject was asked to put his hand on the control . . . but not to take his eyes off the spot (fixation point). After four minutes the top circle was switched 'off', the experimenter's size control set to give six centimetres, and both circles switched 'on' again simultaneously with the instruction to make four post-satiation PSE's.

The test was done in conditions of almost complete darkness. The apparatus was seen by the subject simply as one or two luminous green circles with a small, dim, red spot between them, against a black background.

Brightness Test

A second test, which has been called a figural after-effect of brightness, was included in the test battery. The phenomenon has been called a figural after-effect because although being non-contoured as most visual satiation effects are, it does not differ from them in conditions of presentation or derivation. That is to say, there is no logical difference between the fixation of an inspection figure which is contoured or which is of relative homogeneous brightness. Both figures should produce satiation effects in the central areas mediating their perception and both should affect perception of subsequent stimuli in the same area. This is a deduction which follows from Kohler's original analysis and from the subsequent work of Bevan (1951).

The test takes the simple form of a circle of light composed of two independently variable semi-circles. The subject is asked to make a series of judgements of brightness equality by varying one side of the circle until it is the brightness set by the experimenter. These (pre-satiation) judgements serve as a baseline against which to assess any difference between them and a second series of judgements made after a period of fixation in which only one half of the field is illuminated. Previous work has shown that the second set of judgements of equality are usually determined at a lower level than the originals, i.e. as if the portion of the visual system which had catered for the continuous stimulation and against which the standard intensity is projected, is less responsive than an adjacent area in which the comparison intensity is located.

That these effects in brightness matching are not peripheral is clear from the fact that they transfer from eye to eye, i.e. where one eye is stimulated and the opposite eye is used to make the judgement. The second feature which suggests the classic satiation phenomenon is found in the fact that they only occur when the usual procedure of inspection-test situations are used. (For a discussion of brightness in FAE see Kohler and Wallach (1944), Kohler *et al.* (1952), Hochberg and Triebal (1955), Walthall (1946) and Marquart (1954).

Two important questions may be asked about the present test. (*a*) Is the matching difference due to the characteristic 'fading' of the test figure as it falls upon an area of the retina cortex which has under-gone differential adaptation? and (*b*) Is the effect a contrast phenomenon?

An affirmative answer to the first question cannot be proved from the present experimental method; although it is highly probable that the 'fading' effect mentioned by Kohler and subsequent workers has some effect upon brightness discrimination. Indeed, if the phenomenon is a figural after-effect, fading would be expected, although in the contoured test fading is said to be independent of displacement. The second question can hardly be answered in the affirmative because of the one essential difference between this phenomenon and contrast effects, i.e. time. Contrast effects usually appear immediately or very quickly in a single differential brightness stimulus, whereas the present effect follows from considerable periods of fixation and only after the conditions of prior inspection before test figure have been fulfilled. It cannot, however, be denied that contrast effects, being the product of the degree of excitation will exert some influence, but as Kohler has pointed out, increases in satiation will gradually reduce its effect, and they will be minimal where satiation is greatest.

The prediction in this case is essentially the same as that for the other form of figural after-effect.

Apparatus

The Brightness Test[1] apparatus consists of a double-beam projection lantern having two identical projection systems side by side but deriving their illumination from a single common source. Light is collected from either side of a slightly over-run 36w. 12v. auto-headlamp bulb. Each beam is turned through 90° by identical prisms, to enter duoplano condensers, thereby illuminating evenly the object planes. The objects consist of $\frac{1}{2}$ in high 'D' shaped apertures in thin brass plates, each adjustable laterally by screws against the counter pressure of springs.

Double cell achromatic projection lenses focus the two 'D's back to back to form a circle of light 12 in. dia. 5 ft. distant, upon a white screen. The brightness of either image may be independently varied throughout a wide range by means of slit diaphragms placed between the projection lens units. Colouration is indiscernible until refraction effects between the knife edges occur at the moment of extinction. Mechanical pointers rigidly fixed to the 'shutter' crank-shafts and describing 5-in. arcs against a glass scale, give fair indications of illumination. Each 'shutter' consists of a pair of parallel, straight edged blades in slotted guides moving in opposition. This movement is effected by connecting rods and crankpins set about a disc at 180° to each other. A slow motion drive is fitted to each hand control.

Mechanical back-lash and parallax errors encountered in this measuring system made necessary the use of more refined photoelectric methods for monitoring and light balance determination. Identical thin, clear glass slides were introduced just beyond the projection lenses at such an angle that some of the main beam is scattered on to a selenium cell. The output from each photocell is compared in a bridge circuit and the difference signal passes on to a transistor D.C. amplifier. The associated indicator requires about 300 μA for full-scale deflection and carries a $3\frac{1}{2}$-in. pointer over a dimly illuminated ground glass scale. Graduations are non linear, being opened out about zero. Amplification of about 150 enable input signals of 2 μA to deflect full scale.

---

[1] Acknowledgement is made here for the assistance and advice of S.E.R. Mabel.

Procedure for Brightness Test

In the brightness test the instructions and method were as follows: The subject was seated in front of the apparatus with the occluder removed. He was instructed that the circle of light before him was composed of two equal halves; one which could be varied by the experimenter and one by himself. Demonstrations were given. The subject was then told that, starting from the dark position with his control, he would be expected to match for brightness any value that the experimenter cared to set but that he must keep his eyes fixed upon the small red spot in the centre of the circle exactly as he had done in the previous test. General practice trials were then undertaken. When practice had been satisfactorily undertaken the subject was informed that in the experiment he would have to make his judgements with one eye only as the other would be occluded. After instruction and practice the subject was dark adapted to the standard level of $10^{-4} \times 0{\cdot}60$ c./sq. ft.

When dark adaptation had been accomplished the apparatus was set so that the subject's control was at 'dark' and the standard set at 0·5 lumens although the light beam was shuttered. The subject's right eye was covered. All that he could see at this point was a small red fixation light immediately in front of him. The subject was told: 'I want you to keep your eyes fixed on the red light. When I project a level of brightness on the left-hand side of the circle I want you to match it with your own side using the control you have practised with already. I want you to do it quite quickly but not so as to make errors. Tell me as soon as you think the two sides are equal and then *close your eyes*.'

At this juncture the shutter was withdrawn and the subject made his first match. When the error reading had been taken from the null galvanometer, the occluder was transferred to the left eye and a second trial undergone. These trials were the pre-satiation judgements and they differed only in the levels of illumination set (0·5, 1·0, 1·5, 2·0 lumens).

In the post-satiation judgements the trials were essentially the same, differing only in the following way: 'This time I am going to ask you to look at the small red spot while the left-hand side is illuminated. After one and a half minutes, which I shall call out every thirty seconds, I shall cover the light, transfer the eyeshield to your other eye, and ask you to make another match of equality.'

This was done, the testing being arranged so that the post-satiation judgement was always made with the same eye which had made the pre-satiation judgement. There was a one-minute rest pause, with eyes closed, between pre- and two minutes between post-satiation trials.

No descending trials were undergone because it had been suggested by preliminary experiments that self-satiation was too rapid. To simplify statistical manipulation two scores were derived from the four trials, these are numbers 58 and 59 of the correlation matrix and represent the pooled error scores of the first plus the third and the second plus the fourth trials.

Results

The results of testing for both measures of figural after-effect are presented in Tables 4.5 and 4.6.

TABLE 4.5

Brightness Test
Table showing values and differences for pre- and post-satiation measures at the four levels of illumination. N = 80

	Trials	Pre-satiation Mean	Pre-satiation S.D.	Post-satiation Mean	Post-satiation S.D.	Diff.	*t*	*p*
1.	0·5 lumens	−9·654	6·103	−14·410	8·621	−4·756	4·457	0·01
2.	1·0 ,,	−10·910	7·210	−19·231	11·092	−8·321	7·683	0·01
3.	1·5 ,,	−14·897	8·720	−25·462	12·449	−10·565	8·548	0·01
4.	2·0 ,,	−16·667	9·347	−29·333	14·857	−12·666	8·300	0·01

Readings are in micro-volts taken from deflections from zero on the null galvanometer.

TABLE 4.6

Comparison of Circles
(*Contoured figural after-effect*)
Table showing values and differences for the two ascending and two descending trials. Scores in c.m.

Trials	Pre-satiation Mean	Pre-satiation S.D.	Post-satiation Mean	Post-satiation S.D.	Diff.	*t* =	*p* =
1. Ascending	6·008	0·403	5·471	0·531	0·537	8·262	0·01
2. Descending	6.647	0·505	5·910	0·488	0·764	12·504	0·01
3. Ascending	6·221	0·457	5·814	0·491	0·407	13·567	0·01
4. Descending	6·506	0·474	6·183	0·453	0·323	5·383	0·01
	25·382	1·839	23·378	1·963			
	6·345	0·459	5·844	0·490			

These tables indicate clearly that in both cases the expected after-effects were obtained and are fully significant in all cases. The mean values presented, however, give something of a false picture, and a measurable figural after-effect was not obtained for all subjects on all trials. This is in accord with other findings concerning the universality of the effect. In this investigation a number of subjects failed to observe the effect for individual trials and four of the 80 failed to produce an after-effect at all. None had failed in the pilot experiment where the average decrease in the test figure was 3·02 mm.

The extraversion hypothesis in relation to the magnitude of the after-effect has been examined in the same way as the spiral scores. That is to say, the highest 23 and lowest 23 scorers along the dimension of extraversion, some 29% in each case, have been opposed as separate groups. In the case of the circles test the examination resulted in group means and standard deviations of the combined difference scores on all four trials between pre- and post-satiation scores as follows: Low E, mean 8·696, S.D. = 5·858; high E, mean 7·783, S.D. = 4·888. These scores give a mean difference between the groups of 0·913 which is not significant. The scores themselves suggest that the original hypothesis that extravers should display greater magnitude of after-effects is not supported in this experiment. This failure may lie in the method employed in the elicitation of the phenomenon. One such possible weakness lies in the prior dark adaptation allowing retinal effects to proceed at a faster rate than de-satiation. Another possible weakness may stem from the long period of inspection (satiation). Extraverts are low in persistence and may have found a period of four minutes too long to maintain attention on the fixation point despite the encouragement that was given. Attentional factors cannot be ruled out in this type of visual phenomena and this was one of the reasons why a test of fixation was included in the battery. Performance on the fixation test is correlated with both measures but insignificantly ($r = 0{\cdot}006$ and $-0{\cdot}101$).

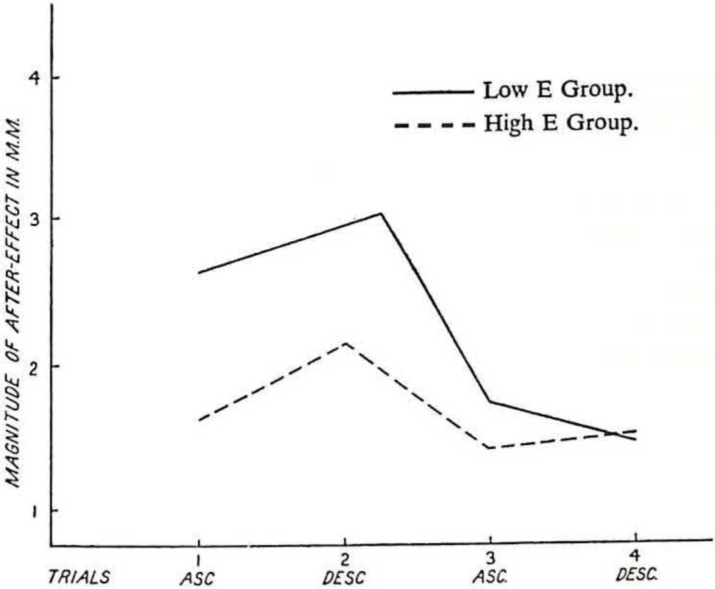


FIG. 4.5. Comparison of Circles (Trials 1–4) Selected Groups.

Another factor which may have resulted in this predictively contrary result lies in the non-transfer method being used. Previous experiments

(Nichols, 1956) had shown that the effect did transfer, indicating that it is a central effect, and therefore, because of test situation difficulties, binocular satiation and testing was conducted and the assumption made that differences in the two techniques would be irrelevant. It now appears possible that this assumption was improper, although experiments using both techniques still require to be done.

Fig. 4.5 displays the relationship between the trial means for the selected high and low extraversion groups (the difference between pre- and post-satiation trials in mm.). It will be seen that there is a large difference in favour of the Low E group on the earlier trials which is not as manifest on the third and fourth trial. Indeed the trend appears to be reversing itself on the fourth trial, suggesting that the persistence of the effect is falling off more rapidly in the Low E than the High E group. (It must be noted here that none of the means differ statistically. Critical ratios = Trial 1: $t = 1{\cdot}68$, $p = 0{\cdot}10$, Trial 2: $t = 1{\cdot}33$, $p =$ n.s.; Trial 3: $t = 0{\cdot}499$, $p =$ n.s.; Trial 4: $t = 0{\cdot}068$, $p =$ n.s.)

The present results are equivocal in relation to Eysenck's statement concerning the magnitude of figural after-effects and extraversion. With a sufficiently large number of trials the observed tendency might have reversed itself to the extent that a mean assessment would be a positive correlate to extraversion. The obtained results, however, suggest that in the early stages, i.e. immediately post-satiation, it is the introvert who has the greater displacement magnitude.

As with other variables, a selected group comparison was attempted for the combined brightness difference scores. Initially the group means obtained seemed incompatible with the positive correlations found between the two threshold measures and extraversion as they indicated that the introverted end of the total group had larger after-effects than the extraverted. However, an examination of the scatter diagram of the combined scores against extraversion indicated lack of homoscedasticity and, therefore, no comparison is given.

The correlational results of these tests are generally equivocal to prediction. There are four measures in all, representing the ascending and descending thresholds on the comparison of circles test, and two thresholds on the brightness test. Of the first test, intercorrelations with the criterion measures of extraversion are generally in the wrong direction from prediction and one of them is almost significantly so, i.e. Circles Ascending and E, i.e. $-0{\cdot}186$. In the brightness test, however, correlations are in the predicted direction, but only one (between the second threshold and MPI (E)) is significant, i.e. $r = 0{\cdot}314$. The first brightness threshold/MPI (E) correlation $= 0{\cdot}147$.

The reliability of the tests cannot be measured directly as there was no retest. However, when assessed by taking the average correlation between trials and correcting for attenuation they are found to be adequate, lying between 0·7 and 0·8.

There are several correlations between these and other measures of specific interest. Two of these have already been mentioned in the section concerned with the spiral scores where it was suggested that figural after-effect magnitude may be a negative correlate of extraversion. Also of interest are the significant and positive correlations between both ascending and descending thresholds on the comparison of circles test and dark adaptation ($r = 0{\cdot}259$ and 0·242). As both correlations are significant, it is unlikely that this is a chance effect and that 'time in the dark', which is common to the two situations, may influence both measures: although this is clear in one case it is not so clear in the other.

Auditory Flutter Fusion

Although an immense amount of research has been undergone to demonstrate the determinants of the visual response to intermittent stimulation and the fusion threshold (Landis, 1954), the investigation of the ear as a temporal analyser where the main variables are similar, Intensity MS. Ratio, etc., has been relatively neglected by psychologists. To some extent this neglect is understandable, for quite recently Stevens and Davis (1938) were writing somewhat contemptuously of the squandering of effort on 'flicker' techniques, claiming that despite the success achieved by such methods in vision, in audition it was a failure. They argued, correctly, that the sudden interruption of a pure tone did not produce a stimulus void but sidebands around the carrier frequency and, consequently, what had been often considered as experiments in auditory flutter were in fact experiments in modulation. The problem of sidebands has been *largely* removed by the use of white-noise, itself a band of all the audible frequencies which when interrupted produces sidebands of each of its component frequencies retaining the uniformity of its spectrum (Miller and Taylor, 1948); Miller, 1947 and 1948; Symmes *et al.*, 1955).

The use of an auditory flutter technique in the field of individual differences (particularly in the abnormal sector (Ax and Colley, 1955; Jones, 1956; Chapman *et al.*, 1955) would appear to hold equal promise with tests of visual analysis. Although the ear is probably not as good an analyser as the eye (Gebhard, 1955; Mowbray, 1955), its range is considerably larger. As Symmes has recently observed, the fusion threshold for light for the vast majority of people occurs at frequencies below 80 c.p.s. whereas, with a mark space ratio of 0·5, auditory flutter may be discriminated at frequencies in the region of 1000 c.p.s. The work of Gebhard *et al.* (1955) supports Symmes' contention in the demonstration that the auditory discrimination range is some 450 times that of vision and that the periodicity ratio is in the region of 16 : 1. Finally, there is, of course, the advantage of a direct energy transformation from light to sound and thus a direct comparison in terms of a common measure of sensory resolution.

The immediate enthusiasm for auditory flutter techniques was dampened by an early adverse report by Goodfellow (1938) who investigated both auditory and tactile thresholds for 'discreteness', demonstrating considerable diurnal variability (3 db. for auditory threshold and 5 db. for tactile) in both functions. The notion of 'discreteness' (and lack of) has also been examined by Ax and Colley who have drawn the interesting distinction between discreteness and flicker. Loss of discreteness, they claim, is a generalized experience not dependent upon the particular sense modality involved. These authors have found high correlations between measures of visual, auditory, tactile and electrical stimulation in 'psychopathological' cases but not in a group designated 'neuropathological', but find insufficient evidence to decide whether temporal acuity is a generalized or specific function of the brain.

One of the most recent and widely accepted hypotheses advanced to 'explain' the phenomenon of auditory flutter fusion has been suggested by Miller (1947 and 1948) and Miller and Taylor (1948). They claim that the cue for flutter is perceived when the 'off' period of the sound/time fraction is longer than the time required for the auditory sensation to decay one detectable step, i.e. one intensity JND. Adducing evidence from their own work (and also of Bekesy (1951) and Symmes (1955)), which has indicated that the auditory decay to threshold is linear, they establish the following equation:

$$f = \frac{\mathrm{I.\ C}}{\Delta\mathrm{I.\ T}} \times 1000$$

where f = the repetition rate for fusion; I = the sound level of the signal in decibels above threshold; T = the duration in milliseconds of the physiological decay to threshold; C = 1·00 minus the sound/time fraction; and ΔI = the intensity JND for noise in decibels. From this equation it is immediately obvious that f is largely dependent upon values of T when I and C are held and ΔI is relatively constant (cf. Licklider, 1951).

Now the reports of Symmes which give distributions of T values, mean = 152 ms. range (104–278 ms.)[1] for a population of 46 cases leads to the reasonable inference that the 'physiological decay' periods are measures of the disinhibitory functions of the mechanisms involved assuming a constant inhibitory effect for a constant stimulus. Schematically this proposition may be shown in Fig. 4.6, where the time ON represents the dissipation rate of a JND (Intensity) for the average person. Within the time ON there will be failure to discriminate flutter whereas at any point beyond N, because more than ΔI has been dissipated, flutter will be observed. The times represented OI and OE are the theoretical positions of introverts and extraverts, differing in the speed at which inhibitory effects disappear.

---

[1] The T values assessed in the present investigation, by assuming a relatively constant ΔI of 0·4 db. have a mean of 242 msecs. and a range from 166–583 msecs. This differs considerably from the values obtained by Symmes and his colleagues and suggests that the sound level meter which was employed to determine the intensity level of the signal was under-reading.

If the above proposition is acceptable, it leads through Eysenck's theory of personality to the immediate prediction of longer decay periods and greater inhibitory effects in extraverts than introverts. It will be remembered that extraverts develop inhibitory potentials more rapidly than introverts and they dissipate these potentials more slowly. It can be deduced, therefore, that auditory flutter thresholds will be higher in extraverts.

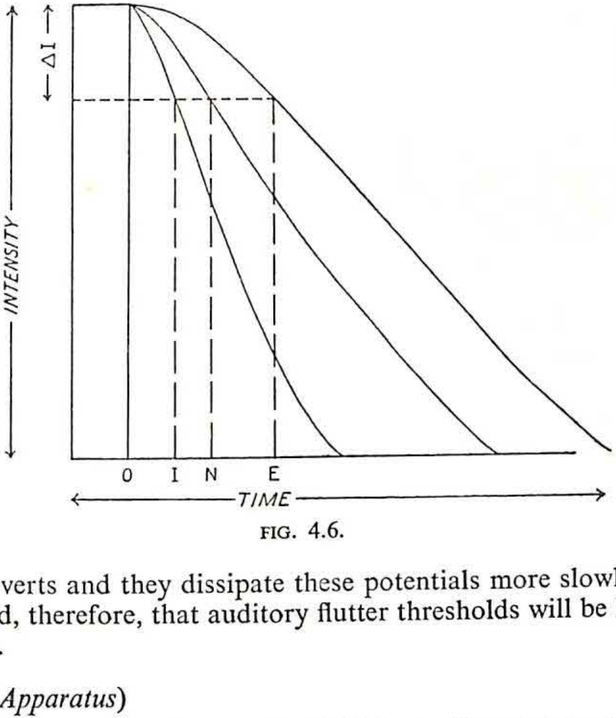


FIG. 4.6.

The Test (Apparatus)

The test used was the University of Chicago Sound Discrimination Test, which has been designed and developed in the Section of Medical Psychology under the direction of W. C. Halstead. Detailed information of the test may be obtained from the manual but the following brief description might be useful. The test is administered by a tape-recorder (in this case a Crofts Radio). The signal consists of interrupted 'white' noise with square wave characteristic at an intensity of 70 db. Repetition rates vary between 10 and 150 i.p.s., in steps of 10, and the sound/time fraction is 0·90. There are four random presentations of the fifteen frequencies by the method of constant stimulus. Instructions, examples, and practice items are all contained in the first part of the recording, and they are reported as being adequate for normal subjects of average intelligence.

Unfortunately, in the present experiment the test could not be given in a sound-proofed room, but all extraneous sounds were dampened by the room preparation which has already been referred to and by 'doughnut' muffs covering the earphones.

Results

The two measures of the test of auditory flutter fusion have been taken representing the threshold (AFF I) and the range of frequencies about

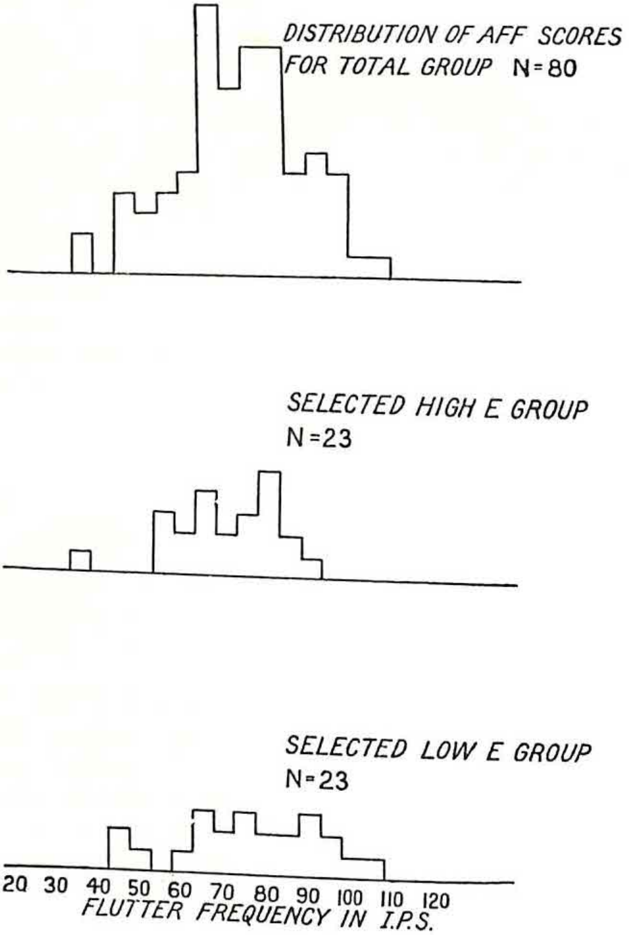


FIG. 4.7.

which uncertainty was expressed by the subject (AFF II), i.e. the range between all right and all wrong. Means and standard deviations were as follows: (AFF I): 72·244, S.D. = 14·586 (AFF II): 45·769, S.D. = 17·980.

Comparison of the two extreme groups on both measures gives the results as shown in Table 4.7.

The means on the first measure are in the predicted direction, i.e. that extraverts have a lower level of discrimination, although the difference between the compared groups of 5·869 c.p.s. is not significant by t test ($t = 1{\cdot}305$, p n.s.). The difference of 0·870 between the groups on the second measure is also insignificant. The results cannot be considered as supporting, at an acceptable level, the initial hypothesis.

TABLE 4.7

Means and standard deviations on the two measures of auditory flutter fusion for the selected groups

	(AFF I)		(AFF II)	
	Mean	S.D.	Mean	S.D.
Low E	75·652	16·73	46·087	17·52
High E	69·783	12·85	45·217	20·64

The correlations of both measures, i.e. threshold and range, are all insignificant with the questionnaire measures of extraversion. The failure of the test measures to correlate with the personality scores is emphasized by the fact that they are not even in a consistent direction, i.e. both coefficients are insignificant but negative with the MPI (E) scale and both insignificant but positive with the CPF (E) scale.

Another feature of interest is the correlation of both measures with the K scale of the MMPI ($r = 0{\cdot}306$ and $0{\cdot}225$) and the D scale of the CPF ($r = 0{\cdot}196$ and $0{\cdot}254$). All four correlations are positive and significant. As one scale (K) is said to measure 'attitude to test items' and the other (D) is essentially a lie scale, it is thought that both measures should be treated with suspicion.

Summary and Conclusions

Six perceptual tests were given to two experimental populations with a total N of 97. These were: dark adaptation, auditory flutter fusion, spiral after-effect and two measures of fig. after-effect. With the tests were administered several personality questionnaires and scales.

Two predictions were made: (*a*) that through the formal and functional identity between 'reactive inhibition', as defined by Eysenck, and 'satiation', as defined by Kohler, a relationship would manifest itself between the perceptual test results and a questionnaire measure of inhibition in the form of 'extraversion', and (*b*) in so far as the tests were indicants of differential inhibition sufficient co-relation should exist between them to suggest that they measured different aspects of a basically similar process.

Neither prediction has been affirmed with any great degree of confidence. Concerning the first prediction only three of the twelve correlations between the MPI scale of 'extraversion' and the experimental variables are significant and one of these is contrary to the hypothesis, as are two others of the remaining nine which do not achieve significance.

The qualification made of the initial hypothesis must also be made of the second. Although a greater number of correlations are significant than could be expected by chance alone, and a large proportion of the remainder are in the predicted direction, sufficient power of correlation does not exist to allow the immediate inference of a common cause.[1] When comparisons are made between two groups representing the upper and lower 29% of the extraversion dimension, there is again general failure to find support for the initial hypothesis.

The measurements taken of the spiral after-effect persistence have emerged as highly reliable despite their failure to distinguish significantly between the selected groups, and the correlational evidence is such as to support the assertion that it is a reasonable measure of extraversion as assessed by the MPI and the C.P.F.

The phenomenon of displacement or sensory reduction following continuous stimulation has been obtained at a satisfactory level of confidence, i.e. despite non-universality, in both measures of figural after-effect. The extraversion/satiation hypothesis, however, is not supported by the comparison of extreme groups or correlation. It has been concluded that the prediction of larger figural after-effects in extraverts is probably incorrect, particularly in the early stages.

The general failure of the two major tests to correlate with, or differentiate between groups based on extraversion is repeated for auditory flutter fusion, fixation, and dark adaptation.

The results of the present investigation are not critical to any theory and support only the complexity of 'inhibition' as an explanatory concept in psychology. Several weaknesses, both methodological and theoretical are obvious and the more prominent may be stated as follows:

(*a*) The measures taken are not related to inhibition or satiation in any straightforward manner;

(*b*) That 'inhibition' is only a secondary characteristic in performance (as measured) instead of the primary which has been postulated;

(*c*) That the tests chosen and administered are not appropriate to the demonstration of measures which can be designated clearly as differential inhibition indicants;

(*d*) That too little is known of the motivational aspects of the behaviour elicited, i.e. attention, interest, instruction, and what Dodge has called 'the unalysed mass of obligations', and the way in which these variables may suppress or change the theoretical course of the inhibitory process;

---

[1] It is, of course, almost impossible to make any assertion as to common causation by a simple visual or item by item examination of such a large matrix, and factorization is more appropriate. The factor analysis of these and other test results will be found in a later section.

(*e*) That the functional identity of reactive inhibition and satiation requires re-examination before the promising line of research which it offers is obscured by premature inferences which fail to allow for difference in 'fatigue' when generated by neural reaction and skeletal response. For instance, the motivational components, which have been already referred to, are usually clear in learning experiments and attempts have been made to lay down the relationships between 'drive' and I_R (fatigue). No such service has been attempted for satiation phenomena although it is known that abnormal groups, which may differ from normals motivationally, also differ in satiation effects;

(*f*) That the personality dimension of extraversion being derived from a questionnaire which is basically behavioural in emphasis fails to account for all aspects of differential fatigue and discrimination in perceptual test measures.

Once these weaknesses have been removed and the functions of satiation clarified, the hypotheses which have been examined in this investigation will, perhaps, emerge in a more clearly testable form.

5

A FACTOR ANALYSIS OF SELECTED TESTS

H. J. Eyesnck

THE design of the experiment, details of which have been reported in the two previous sections, was geared to a factorial analysis of the intercorrelations between the tests, and we must now turn to such an analysis. Fifty-nine scores in all were intercorrelated but only thirty-four of these were selected for a factor analysis. The reasons for the exclusion of the other scores lay mainly in the fact that they were alternative forms of the same test, or else scores of part of a test where only the total score was used in the factor analysis, and other reasons of a similar nature. The actual thirty-four test scores used are given in Table 5.1. They are not described in detail because the reader will find the detailed descriptions in the previous two sections; enough is said under the heading 'Nature of Test' to give a fair indication of the nature of each score without having to turn back to a previous page.

The correlations were subjected to an analysis along the lines of Hotelling's method of principal components; the work was done on an electronic computor under the supervision of Dr. P. Slater. Three factors take up almost all the communal variance, and the loadings of these three factors, rotated according to the dictates of Thurstone's principle of simple structure, are given in Table 5.1. (Five factors in all were taken out, but two were disregarded because of their low loadings.)

We may first of all dismiss the third factor which appears to be an artefact created by the presence in the battery of several scores derived from the *spatial conditioning* and the *verbal conditioning* measures. Such scores are more of the nature of reliability coefficients, and it follows that the loadings on this factor are relatively high but have little if any psychological interest. It will not be discussed further in this chapter.

Factors 1 and 2, in accordance with the hypothesis underlying the selection of tests, should be extraversion and neuroticism. The reader may find it advantageous to consult, in trying to answer the question of whether the results bear out the hypothesis, Fig. 4.1 (p. 195), which presents the factor loadings on Factors 1 and 2 in diagrammatic form. Lines have been drawn at values of +0·25 and −0·25 as a rough visual check on the adequacy of the simple structure solution; it will be seen that hardly any tests lie outside the centre cross, and those that do are only very little removed from the inside.

We must first of all enquire regarding the position of the various questionnaires included, as previous experience with them gives us a fairly clear picture of the kind of solution expected in terms of our hypothesis. It will be seen that Factor 1 is characterized by the MPI extraversion score (−0·540), the Cattell extraversion score (−0·452), the MMPI psychopathic deviate scale (−0·458), and the MMPI hysteria scale (−0·307). These are the four scales expected to be measures of extraversion, and all of them are found to have medium to high loadings on Factor 1. We may conclude that in terms of questionnaire measurements Factor 1 may be identified as one of extraversion. (It will be noted that all the loadings have a *negative* sign, so that it is the negative end of the factor which may be regarded as the extraverted end, while the positive end is to be regarded as the introverted end.)

It is interesting to note that both the Cattell and the MPI extraversion scales have nearly identical loadings on this factor, because both were developed quite independently, and indeed until quite recently Cattell was rather critical of the very notion of extraversion as a factor. The MMPI scales, which were developed independently of both the Cattell and the MPI scales also have comparable loadings, as indeed might have been expected on the hypothesis that both hysteria and psychopathy are *extraverted* neurotic disorders. The result would seem to provide some evidence of validity for this hypothesis, and also for the validity of the MMPI scales. It will also be noted that in all the analyses reported to date (Hildebrand, Claridge, Willett-Holland) the extraversion questionnaire used had a loading of about 0·55 (0·563; 0·48, 0·64; and 0·540 respectively). This constancy in spite of differing samples of subjects and tests is encouraging. The actual values are rather low, however; they suggest that only something like 30% of the total variance of the factor is being measured by this scale. This puts severe limits to its use as a *criterion*; to obtain an estimate of the correlation of any given test with the *factor* of extraversion, when only its correlation with the questionnaire (r_q) is available, the most likely true correlation r_t would be:

$$r_t = \sqrt{3r_q^2}$$

17	42	Verbal Conditioning: Overall difference score: B–2A	0·071	−0·298	0·511
18	43	,, ,, Slope	0·177	0·149	0·345
19	44	,, ,, Rank mean square deviation variability about previous score	−0·165	0·227	−0·441
20	45	IQ (NIIP Test 90)	0·010	0·224	−0·033
21	46	MMPI HY	−0·307	0·373	−0·210
22	47	PT	−0·058	−0·633	0·036
23	48	Pd	−0·458	−0·142	−0·114
24	49	K	0·009	0·753	0·008
25	50	CPF Extraversion	−0·452	0·137	−0·020
26	51	Lie scale	0·027	0·687	0·236
27	52	MPI 'E'	−0·540	−0·049	−0·011
28	53	'N'	0·078	−0·582	−0·063
29	54	Fixation. (Failure to maintain F)	−0·026	−0·400	−0·066
30	55	Figural After-Effect—circle. Descending trials	0·460	0·065	0·171
31	56	,, ,, ,, Ascending trials	0·551	0·033	−0·014
32	57	Spiral After-Effect. ε of two 100″ trials of prior stimulation	0·726	−0·164	−0·069
33	58	Figural After-Effect: Brightness. 1 + 3 pre- 5 + 7 post	−0·068	0·169	0·094
34	59	,, ,, ,, 2 + 4 pre- 6 + 8 post	−0·096	0·164	−0·088

The usual run of r_q correlation is around 0·3, which would give an r_t correlation of above 0·5. These values, in turn, would of course require to be corrected for attenuation in the experimental test which was being correlated with the questionnaire, particularly when the reliability of such a test happened to be low. Some further discussion of this point is given in the Epilogue.

The second factor has its highest loadings on the MMPI psychasthenia scale (−0·633), the MPI neuroticism scale (−0·582), the MMPI K-scale (0·753), and the Cattell lie scale (0·687). These four scales define the factor very accurately, all of them lying almost exactly on one line, and strongly support the hypothesis of Factor 2 as a neuroticism factor, at least as far as its verbal components are concerned. We may note with some surprise that the hysteria scale has a *positive* loading on this factor, i.e. is situated towards the normal rather than the neurotic end. It is difficult to know whether this loading is significant, and if so whether it would be duplicated in research using a less extraverted sample of subjects. It should also be noted that the MPI extraversion and neuroticism scales are located in positions such that the angle of the vectors joining them to the origin is almost exactly 90°. This agrees well with the usually observed independence of extraversion and neuroticism measures in normal samples. There is here a considerable difference from the results of the Claridge study, and we will return to this later.

Having decided provisionally that Factors 1 and 2 may be identified with extraversion and neuroticism respectively, we must now turn to consideration of the various scores derived from objective tests, in order to see to what extent these conform with prediction. The first test to attract attention by virtue of its high saturation is the length of after-effect produced by a rotating spiral. Four scores are included in the analysis of which two are direct measures of the length of after-effect, namely No. 32 and No. 4. These are expected to correlate positively with introversion and should therefore have positive loadings on Factor 1; the loadings actually found are 0·726 and 0·685. It may be concluded that the prediction is verified. As regards the two other scores derived from the spiral, namely the 'slope' and 'distance' scores, it is difficult to interpret these or derive any special predictions for them. It might be argued that in terms of the theory outlined in H. Holland's section, a high 'distance score' should be found in extraverts by virtue of the fact that the inhibition which is supposed to underlie the curtailment of after-effects would be more effective with extraverts at longer periods of rotation, thus producing a negatively accelerated curve, and the fact that the score in question has a negative loading (−0·222) would seem to support such an interpretation. However, the loading is too low and the prediction too uncertain to inspire much confidence.

The next test to be considered is the spatial conditioning test. Here, according to the hypothesis, we would expect scores 15 and 16 to correlate positively with introversion, and indeed both scores have positive loadings on Factor 1 (0·256 and 0·446). The prediction is therefore verified, but it should be remembered that the status of the spatial conditioning test as a measure of *conditioning* is doubtful for the reasons indicated by R. Willett in his section. Clearly further work is required on this test before we can confidently regard it as a true measure of spatial conditioning, and until such investigations have been carried out it is impossible to come to any firm conclusion about the apparent verification of our prediction.

We next turn to the two measures of figural after-effect included in this study. The prediction linking length of after-effect with extraversion is clearly not verified. The test using contours (circles) has quite high loadings on introversion (0·460 and 0·551 respectively for descending and ascending trials). The scores of the test involving brightness are in the predicted directions, but the loadings are too low to be of any great interest (−0·068 and −0·096).

Before regarding these results as a disproof of the hypothesis, we may consider the argument put forward in the epilogue, according to which we have two possible effects produced by the inhibitory potential underlying extraverted behaviour. In the first place we have the fact that inhibition should mediate strong figural after-effects, a prediction clearly not verified in this study. In the second place, however, we have the prediction that inhibition should produce difficulties in maintaining fixation over long periods, an effect which in view of the importance of fixation for the occurrence of figural after-effects would lead to a lesser degree of figural after-effect in extraverts. We thus have two opposing hypotheses, and it seems necessary to introduce as an important experimental parameter the actual length of stimulation. Figural after-effects can be produced by quite short periods of stimulation, but it is unlikely that difficulties in maintaining fixation would arise until some considerable time had elapsed. It might therefore be hypothesized that *with short periods* of inspection satiation would be generated more strongly in extraverts, as predicted, and the process would not be disrupted by difficulties in maintaining fixation. This might be true with periods up to one minute, say. After that point we might expect difficulties in fixation to arise and more or less cancel out the strong satiation effects produced by extraversion. With periods of three or four minutes we might expect difficulties in fixation to more than outbalance the stronger satiation developed by extraverts, and actually lead to an inversion of the prediction, so that with very long periods we may find introverts showing greater figural after-effects.

The facts seem to support some such interpretation. Rechtschaffen (1958), using a forty-second period of fixation, obtained results showing stronger satiation for extraverts at a level not far removed from significance on the appropriate one-tail test. Nichols (1956), using periods of two minutes of fixation, and Holland in the brightness figural after-effects study included in this factor analysis, using ninety seconds of fixation, both obtained results showing no marked difference either way between extraverts and introverts. Holland, in the other figural after-effects study included in this factor analysis, used a four-minute period of inspection and found a quite marked relationship between introversion and figural after-effects. While it is true that these writers all used different tests and different types of subjects in their various studies, it nevertheless remains true that the results do fall into a clear-cut progression such that short periods of inspection appear to indicate *greater* figural after-effects in extraverts, while long inspection periods appear to show less figural after-effects in extraverts, with intermediate periods of inspection giving neutral results. This would appear to be a hypothesis well worth testing directly in future research.

It will be noted that test No. 29, which is labelled a test of failure to maintain fixation, is loaded on neuroticism (−0·400) rather than on extraversion (−0·026). This would appear to contradict our hypothesis, but it should be pointed out that the interpretation of this measure as one of failure to maintain fixation is by no means well-established, and requires a direct and formal proof before being acceptable. Even if it should turn out that this measure was indeed what it purports to be, it would still be possible to modify the hypothesis outlined above by postulating some other inhibitory effect increasing in time and counteracting satiation. There are several candidates for such a position. With maintained inspection the inspection figure tends to become less vivid and may occasionally disappear completely; this type of satiation effect, which according to the hypothesis would be stronger in extraverts, might serve to reduce the impressiveness of the inspection figure and thus reduce the strength of the after-effects produced by it. (Cf. the first part of the hypothesis relating to spiral after-effects, where such an effect is explicitly postulated.) Other possibilities of a similar kind will no doubt occur to the reader, and we must conclude that the data reported here, while suggestive, do not by any means make a final judgement on our theory possible.

We may next turn to the score on dark adaptation, which has a loading on extraversion (−0·293 and also a loading on neuroticism (0·254), which would indicate a tendency for the more neurotic to have better dark adaptation than the more normal. The first of these findings, if it could be regarded as significant, would be in line with the general tenor of satiation theory. According to this theory the constant bombardment of the retina during the day by light produces a state of satiation stronger in extraverts than in introverts and needing longer periods of dark adaptation to produce an equal degree of sensitivity. It is not known, of course, whether this argument is in fact true, and a more direct test of the theory might be desirable. Nevertheless, as far as it goes, the argument would seem to lead to the prediction that dark adaptation measures of the particular kind here used would show a poorer dark adaptation in extraverts, as is indeed found. The failure of neuroticism to correlate positively with poor dark adaptation is less easy to understand as in the past results have usually tended in that direction (Granger, 1953; 1954, 1955, 1957). However, it is well known that dark adaptation is a highly complex phenomenon and correlations can be expected to vary according to the precise type of apparatus used and the precise method of measurement chosen.

Both verbal conditioning (0·071) and eye-blink conditioning (0·116 and 0·177) were expected to correlate with introversion, and while the factor loadings are in the right direction, they are clearly too small to carry conviction. While the prediction regarding the verbal conditioning score was predicated on the hypothesis that this was indeed a measure of conditioning, a hypothesis for which there is no strong evidence, the same excuse cannot be made for eye-blink conditioning, which had on previous occasions correlated quite highly with extraversion. The only possible explanation of its failure to produce a high factor loading on introversion can be an appeal to sampling errors. The argument has already been presented by R. Willett and will not be repeated here. It would obviously be desirable for the experiment to be repeated in order to see whether this possibility can indeed be admitted as a likely explanation. There is a slight tendency, as shown by score 19, for verbal conditioning scores to be more variable among extraverts (−0·165), and this would seem to be in line with the general evidence of greater change and performance by extraverts; however the loading is too low to allow of any particular interpretation.

We must now turn to the tests correlated with neuroticism. Mention has already been made of test 29, or failure to maintain fixation; (−0·400); until the nature of this test is made a little more explicit, however, no reasonable interpretation of it can be given at the moment.

Much the same is true of tests 9 and 10, both of which are measures of salivation (−0·468 and −0·539). This test had originally been intended as a measure of conditioning, and as such it would have been expected to correlate with introversion rather than with neuroticism. However, as shown by R. Willett, supplementary experiments have thrown considerable doubt on this interpretation of the test, and it seems likely that it must be regarded largely as the effect on salivation of a simple light stimulus. Why such a stimulus should produce effects of this kind is not obvious at the moment, and until more work has been done on salivatory conditioning, so-called, it is impossible to interpret the result.

Lastly we come to tests 2 and 3, which are measures of auditory flutter fusion (0·367 and 0·309). It would appear that neurotics have a lower flutter fusion threshold, and it would also appear that there is no relationship between these thresholds and extraversion. The results seem to be in good agreement with the work of Krugman (1947) and Goldstone (1955) who, working with visual flicker, both found that the more neurotic groups tended to have lower thresholds while the more normal groups tended to have higher thresholds. The position is rather complicated by the fact that both Krugman and Goldstone used dysthymics as representatives of their 'high neuroticism' groups, thus leaving it open whether the low CFF thresholds were due to the neuroticism or the introversion of the subjects. In the *Dynamics of Anxiety and Hysteria* I put forward a theory linking CFF to the extraversion-introversion continuum, predicting '*in extraverts and hysterics lower thresholds of temporal discrimination*'. Evidence on this point is not conclusive, but if it were true, as the present study suggests, that neurotics are characterized by low thresholds of temporal discrimination, then the hypothesis linking flicker and fusion thresholds to extraversion would have to be reconsidered and might in fact have to be inverted.

This concludes our discussion of the results of this factorial study. It may be of some interest to compare some of the results with those reported by G. Claridge. It is by no means obvious that results, even of similar tests, should maintain similar relationships in two groups differing as much as those here considered. As an example of the kind of difference to be expected, let us consider the position of the two MPI scales. It will be remembered that extraversion and neuroticism are quite unrelated in the Willett-Holland study, extraversion having a loading of −0·049 on the neuroticism factor, and neuroticism having a loading of 0·078 on the introversion factor. When we come to the Claridge study we see that this is by no means true. Here extraversion has a loading of −0·53 on the neuroticism factor, while neuroticism has one of 0·09 on the extraversion factor. In other words, among neurotics extraversion has a strong negative loading on neuroticism, while among normals it has quite a negligible loading on neuroticism. This fact had been observed several times before and indeed has been remarked upon in the Manual of the MPI (Eysenck, 1959b). Clearly there are interaction patterns at high neuroticism and high introversion scores which do not permit us to retain the hypothesis of linear regression. It is not impossible that other similar effects will be found in connection with other tests, and it would seem that an investigation on a comparative basis of intercorrelations of different test scores in a normal and in a neurotic group would be of considerable interest and importance.

Another test which clearly differs in its factorial orientation in the two studies is the spiral after-effect. In the Willett-Holland study spiral after-effect is highly correlated with introversion and very little with neuroticism; in the Claridge study it is highly correlated with neuroticism and only slightly with introversion. In both studies, therefore, it is the dysthymic who comes out as having the strongest after-effects, but the main stress in the one case is on neuroticism, on the other introversion. Why this should be so, is by no means clear.

Auditory flutter fusion also does not give identical results in the two studies. In the Willett-Holland study it will be remembered that auditory flutter fusion had a negative loading on neuroticism; this is true also in the Claridge study, but only to the quite insignificant extent of a loading of −0·07. Auditory flutter fusion, however, has a loading of 0·39 with extraversion in the Claridge study, while it has no such loading in the Willett-Holland study.

The only remaining tests which occur in both batteries are the MMPI scales, and these again provide considerable contrast. All the scales used have very high loadings on neuroticism in the Claridge study and no loadings at all on extraversion; it will be remembered that in the Willett-Holland study both the hysteria and the psychopathic deviate scale had high loadings on extraversion but not on neuroticism. These differences, of course, are understandable in terms of the much greater variance on neuroticism in the group studied by Claridge as compared with that studied by Willett and Holland; if corrections were made for this difference in variance, then rather more similar results might be obtained.

It seemed worth while to draw attention to these comparisons in spite of the small number of tests which was common to the two batteries, because it seems to be likely on methodological grounds that in future work much useful information could be gathered if identical batteries of tests were administered to groups of subjects widely differing in respect of their positions on dimensions under investigation, while as similar as possible with respect to irrelevant dimensions (age, sex, social background, intelligence, etc).[1] It is easy to predict, and even to control statistically, the simple effects of differences in means and variances along the dimensions involved, i.e. neuroticism and extraversion in this case; what is of much greater interest, and indeed constitutes an almost virginal field of exploration, is the problem of non-linear regressions, interaction of factors at extremes, and so forth. Such findings as those relating to the interaction between introversion and neuroticism demand a theoretical explanation and are likely in turn to lead to interesting and important experiments which might throw light even on problems other than those directly involved.

---

[1] In our experience, it has usually been easier to verify hypotheses when dealing with *neurotic* than with *normal* subjects. Similarly, it has usually been easier to verify hypotheses with older than with adolescent (or even younger) subjects. The latter finding may be due to the relatively unstructured personality of young subjects (cf. Eysenck, 1960). The former finding may be due to the multiplicative action of D (anxiety), which, being stronger in neurotics, produces greater differences between extraverts and introverts where these differ in any characteristic which is multiplied by D to produce $_sE_R$. Both explanations are speculative, but it should not be impossible to test them experimentally, or to investigate the correctness or otherwise of the impression mentioned above concerning different groups. The greater success of the Claridge experiment, as compared with the Willett-Holland study, may be due to the fact that in point of time his experiment was planned *after* the other one had been completed, thus enabling him to benefit from the experience of Holland and Willett.

SUMMARY

In this chapter the factor-analytic study has been reported of a variety of verbal and performançe tests of personality. These were selected in the hope of providing measures for the two factors of extraversion and neuroticism, and the factor analysis has clearly demonstrated that this hope was justified. A large number of individual predictions was made and specific hypotheses tested; some of these were verified, some failed to be verified; others had to be modified, and in others certain parameters had to be added to the statement of the hypothesis to incorporate the results. A comparison was made of the results of the analysis with those of another study carried out on a group of neurotic subjects of similar age, sex, social class and intelligence, and certain interesting and instructive differences observed.

PART THREE
Experiments in Psychometrics

S.C.E.R.T., West Bengal
Library
Calcutta
25/3, B. C. Road

6

FACTOR ANALYSIS AND SOME ALLIED PROCEDURES

P. Slater

---

1. THE GEOMETRICAL MODEL USED IN PRINCIPAL COMPONENT ANALYSIS

THE methods of analysis to be discussed all apply to data which can be set out in a table with a number of columns, say *m*, and a number of rows, say N. Each column refers to a psychological test and each row to a person: the entry or element in the Nth row and *m*th column is the Nth person's score on the *m*th test. Any such N × *m* table can be represented geometrically by a dispersion or scatter of N points in an *m*-dimensional space. To examine this geometrical model let us begin with one person's score on two tests.

As a 2-dimensional space we may use any plane surface, e.g. the surface of a piece of paper. Any position on such a surface can be specified by a pair of measurements if two intersecting lines or axes are laid down and its distance from each of them is measured; and conversely any pair of measurements specifies one position. When the axes are used to represent the two tests the pair of scores the individual has obtained locates a position for him on the plane. Thus in Fig. 6.1 the position P is referred to the axes OX and OY by the measurements +1·4, +2·0. They indicate that a movable point will arrive at P from O by travelling 1·4 units of distance in the direction of X and then 2·0 units in the direction of Y, i.e. parallel to the axis OY. In imagining them as carrying the point from O to P we are treating the pair of measurements as a vector. Another vector, −1·8, +3·0, would carry the point from O to a different position, Q (the negative sign serves to indicate that movement parallel to OX is in the direction away from X).

Any such vector, i.e. the co-ordinates giving the position of a point anywhere in the 2-space, can be used to define an axis. For instance passing a line through O and Q and continuing it indefinitely in either direction defines the axis OQ. There is no limit to the number of axes that can be laid out on the same surface with the same origin. Any two vectors, provided they do not coincide, will serve as reference axes for locating a point on it. But a particular pair may be particularly convenient for a particular purpose. Thus in Fig. 6.1, OQ might be considered preferable to OY for use in conjunction with OX, because OQ and OX are at right angles to one another. In other words, OY projects onto OX, OQ does not.

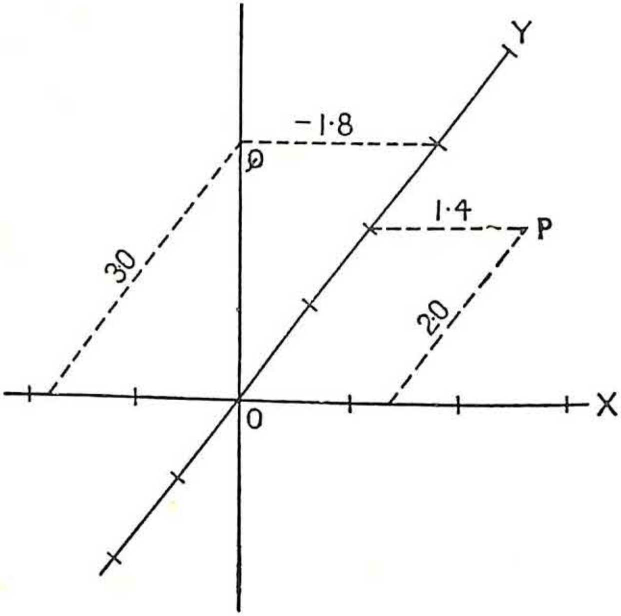


FIG. 6.1.

But we do not require all the possible positions a point can occupy in the 2-space to specify all the possible dimensions or axes. The points lying on the circumference of a circle with its centre at O are sufficient. It is customary to use a circle with unit radius for this purpose. When the reference axes OX and OY are set at right angles to one another as in Fig. 6·2, the co-ordinates of any such point satisfy the condition $x^2 + y^2 = 1$; they constitute what is generally called a normalized vector.

This model can be extended to represent the results obtained when N people have been given two tests, X and Y, and their scores have been standardized (means equated to 0 and s.d.s to 1). The reference axes OX and OY are set at right angles to one another on a plane, cf. Fig. 6.2, and the pair of scores obtained by each individual is used to locate his position. Evidently a scatter of N points will be obtained; and as the tests usually exhibit some degree of correlation it tends to be elliptical, as in Fig. 6.2, where it is shown by a contour line and a correlation of 0·5 is illustrated.

This ellipse has two canonical axes, the major one OP and the minor OQ, defined respectively by the vectors +1, +1 and +1, −1, which are its latent vectors and may be denoted l'_1 and l'_2. A factor of $1/\sqrt{2}$ can be introduced to normalize them. The first defines the dimension in which variation is greatest and the second the one in which it is least.

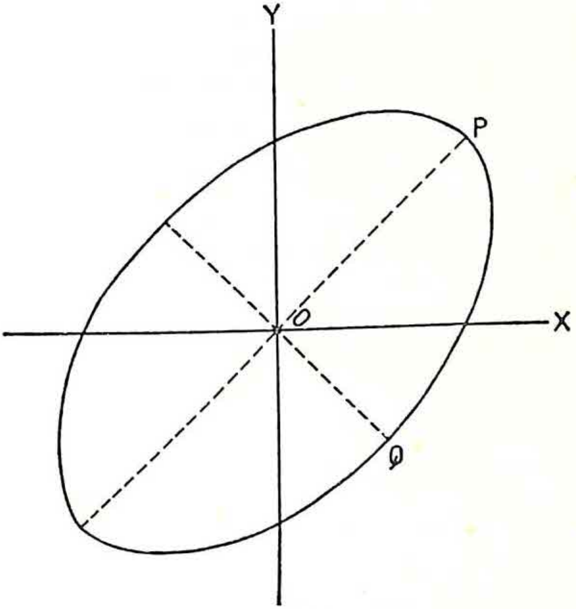


FIG. 6.2.

The equation for the ellipse takes its simplest form when the ellipse is referred to its canonical axes, and this is why the term canonical is applied to them. The equation then becomes

$$\frac{p^2}{\lambda_1} + \frac{q^2}{\lambda_2} = 1$$

Here λ_1, is the square of the distance from the centre of the ellipse to its circumference along the major axis and λ_2 is the square of the distance along the minor axis; the λ's are known as the latent roots. (Note that when $\lambda_1 = \lambda_2$ the ellipse is a circle.) The position of a point given as x_i, y_i with reference to the X and Y axes can be specified with reference to the P and Q axes by using the l's as weights in a weighted summation. Thus $p_i = (x_i + y_i)/\sqrt{2}$ and $q_i = (x_i - y_i)/\sqrt{2}$.

When the scores on the two tests are left unstandardized their scatter still remains elliptical but the centre of the ellipse may be at any distance from the intersection of the two reference axes, and the canonical axes may intersect the reference axes at any angle. Consequently the equation of the ellipse is much more cumbersome and in general its properties are much less easily described. However, there may be some conditions under which standardization should be avoided. The assumption that the N people differ from each other to just the same extent in one psychological characteristic measured by X as in another measured by Y cannot be exactly correct and in many circumstances it may not even be plausible, cf. Thomson (1950). It is excusable only as a first approximation when accurate information is lacking.

When the scores are standardized but left referred to their original axes (X and Y in Fig. 6.2) the equation takes the form

$$x^2 - 2\,rxy + y^2 = k$$

where r is the correlation between the scores in X and Y and k is an arbitrary constant. Taking different values of k will give a set of concentric ellipses with the same ratio $\lambda_2 : \lambda_1$. Thus the ellipse encompassing the area within 1 s.d. of the bivariate mean point is specified by $k = 2(1 - r^2)$.

This equation appears quite tractable, for as there are only two tests only one cross-product term appears in it. But when the number of psychological tests used in the experiment increases to m, and the geometrical model expands from a 2-space to an m-space, the number of cross-product terms in the equation for the ellipse, standardized but left referred to its original axes, increases to $m\ (m - 1)/2$; and the mathematical convenience of using the canonical form becomes so much the greater. To describe how it can be put in this form I shall use matrix notation, and must apologize for introducing it without any adequate explanation for lack of space.

Readers unfamiliar with the notation will be able to keep track of the argument, I hope, if they bear in mind that matrix algebra enables processes of addition, subtraction, multiplication, etc., to be applied to orderly arrays of numbers, such as vectors, which consist of single rows or columns, or larger matrices which may have any number of rows and columns. Any such matrix may be denoted by a single symbol; for instance R is used in the next paragraph to denote a table of correlations between m variables. The position of an element, i.e. a particular entry in a matrix is given by two subscripts, the first specifying the row in which it occurs, the second the column. A matrix is transposed when the rows are re-written as columns; it is unaffected by transposition only if it is symmetrical about its leading diagonal. Thus we could write

$$\mathrm{R} \equiv \mathrm{R}'$$

meaning that R is identical with its transpose. For the typical element in r, namely r_{ij}, the correlation between test i and test j, is identical with r_{ji}—in other words, it is entered in row i column j and again in row j column i.

The m variables may be denoted by a column of x's with tranpose $x' = (x_1, x_2 \ldots x_m)$ and the correlations between them may be written as

$$R = \begin{bmatrix} 1 & r_{12} & r_{13} & .. & r_{1m} \\ r_{21} & 1 & r_{23} & .. & r_{2m} \\ r_{31} & r_{32} & 1 & .. & r_{3m} \\ .. & .. & .. & .. & .. \\ r_{m1} & r_{m2} & r_{m3} & .. & 1 \end{bmatrix}$$

i.e. as a symmetrical square matrix with units in its diagonal. The equation for an ellipse describing the multivariate frequency distribution can then be given as

$$x'R^{-1}x = k$$

where k is an arbitrary scalar quantity as before. Differentiating we find that the particular values of x, sc. $x = l$, which define maxima and minima are specified by the 'characteristic equation'

$$Rx = \lambda x$$

where λ enters as a Lagrange multiplier. This equation can be understood as implying that at points such as P and Q in Fig. 6.2 where the distance from the circumference of the ellipsoid to its centre is locally maximum or minimum the line perpendicular to the tangent (or the tangential hyperplane in the m-space) is continuous with the line from the circumference to the centre, so both are defined by the same vector. Looking at the Figure it is easily seen that this is a property which must characterize the maxima and minima, and that other points on the ellipse do not possess it. But the points on the circumference of a circle or hypersphere all exhibit it, for all are equidistant from the centre.

The best known method for obtaining the numerical solutions of the characteristic equation for a given R is the one due to Hotelling, which is fully explained by Thomson (1950), Holzinger and Harman (1941), and others. But very powerful methods have been developed more recently for use with electronic computers: they should certainly be employed where the facilities are accessible. The resulting λ's and their associated vectors, the l's, are often known as the principal components of the matrix of correlations.

The measurements on a λ scale of an individual whose standardized scores on the m tests are specified by the row vector x'_i, can be obtained by computing $1'x_i$. For this purpose the elements of l must be normalized, i.e. scaled down to satisfy the condition $l'l = 1$.

The first latent vector, l_1, defines the dimension in the m-space where variation is greatest, and its associated root, λ_1, defines the extent of variation in that dimension. The successive latent vectors $l = l_1, l_2 \ldots l_m$ define dimensions at right angles to one another in the original reference space and where variation is uncorrelated and progressively smaller in extent. These properties can be summarized conveniently when the subscripts j and k are introduced to specify two non-zero roots and their associated vectors.

1. The latent roots are progressively smaller, viz.:

$$\lambda_1 > \lambda_2 \ldots > \lambda_m$$

2. The vectors are at right angles, viz.:

$$l'_j l_k = 0$$

3. They are uncorrelated, viz.:

$$l'_j \mathrm{R} l_k = 0$$

2. WEIGHTS USED IN PARTIAL REGRESSION EQUATIONS AND AS DISCRIMINANT FUNCTIONS

In many psychological experiments the m tests are used as a battery in the hope that some linear combination, i.e. weighted summation of them will provide a reliable estimate of a criterion. For instance a battery of tests might be tried out in developing a procedure for selecting salesmen. It could be administered to a relatively unselected sample of persons who afterwards pass through a standard training or probationary period. A careful assessment of their success is made; it constitutes the criterion, Y. Correlations between the tests and the criterion, say

$$y' = (r_{1v}, r_{2v}, \ldots r_{mv})$$

could be computed and the multiple regression equation

$$\mathrm{R}x = y$$

solved to obtain the particular set of values of x, usually denoted $x = \beta$ and described as the partial regression coefficients, by which Y can be estimated with minimum error. A battery of tests of general utility could be related to an unlimited number of criteria in this way, and for each criterion a somewhat different set of β's and a different amount of residual error would be found. The expected value of Y for an individual with a given set of standardized test scores, x_i, would be obtained as

$$y_i = \beta' x_i$$

and would provide the basis for a decision about his disposal.

In other investigations the m tests may be used experimentally for discriminating between members of two contrasted groups. We might for instance contrast people susceptible to an infection with people immune. An appropriate set of measurements could be obtained from sample of persons all subsequently exposed to the infection, and the differences between the means of the two groups found as

$$d' = (d_1, d_2 \ldots d_m)$$

The analogous equation

$$\mathrm{R}x = d$$

is solved to obtain another particular set of values of x, say $x = \gamma$, generally known as discriminant functions. When substituted in the equation

$$d_i = \gamma' x_i$$

they will give a discriminant score for an individual from which he can be allocated to one group or the other with a minimum risk of error.

3. INTERNAL FACTOR ANALYSIS

All such β's and γ's are vectors defining dimensions in the m-space which may be of psychological or other interest in particular contexts. Unless it relates to some external criterion of specific interest there might not seem to be any reasons other than those of mathematical expediency for preferring one vector to another; all might appear equally worthy of consideration. The fact is, however, that psychologists are often predisposed in favour of particular vectors and can give reasons for believing that some define primary, unitary or fundamental psychological traits while others do not.

To avoid the use of affectively toned words, let us suppose that F1 and F2 are two such traits which cannot be assessed directly. They are considered to be independent of one another. Individual differences in F1 are believed to affect scores in psychological tests of one kind, the reference tests for F1; differences in F2 affect scores on other tests, the reference tests for F2; and scores in some further tests may be affected by both F1 and F2 to differing, unknown extents. The m tests are chosen to include some reference tests of both kinds as well as the tests *sub judice*. Tests suspected of assessing other characteristics in common are avoided.

If the results are in accordance with expectation, substantial correlations will be found among the F1 tests and among the F2 tests. Neither kind should correlate with the other, but both should have some correlations with the test *sub judice*. Two of the latent roots of R will be relatively large, the remainder will all be small and may not vary widely. The vectors associated with the first two roots will define a sub-space of two dimensions in the m-space which is called, when generalized, the common factor space, or f-space for short, the residual sub-space being the specific factor space or s-space.

But they are not the only vectors which will serve to define the f-space. Any other pair of vectors lying in the same plane will serve equally well, cf. Fig. 6.1 and the accompanying discussion. Techniques for rotating vectors are well known (Fruchter, 1954, for example) and they can be extended to any number of dimensions. So if there are valid psychological grounds for preferring a particular pair there are no mathematical objections. And again if equally cogent psychological arguments can be adduced in favour of another pair the mathematician will be well advised not to attempt to arbitrate. In spite of such hazards a consensus of opinion tends to grow up in course of time about the tests which should be used as reference tests for certain factors, and as it comes into existence the allocation of tests still *sub judice* tends to become less controversial.

The unrotated factor loadings are not always obtained as the l's in a principal component analysis. Other methods are employed. Consequently the f-space may not always enclose the maximum possible part of the m-dimensional dispersion, and the s-space will then retain rather more than it need. It will be still further inflated if one or two factors are discarded after rotation. But the outcome of these incomplete analyses is always to separate a residual sub-space with relatively small canonical axes from the common factor space, where the canonical axes are relatively large. The f-space is treated as containing the variation of psychological interest and the s-space as psychologically negligible.

Even if it has acquired a psychological reputation beyond the context of a single experiment a factor is defined within that context by a set of loadings denoting a vector in the f-space enclosed within the m-space. From this point of view it must be regarded as a linear dependent function, i.e. a weighted sum of the scores on the original battery of tests. It cannot be equated with any other vector or weighted sum of score, obtained from another independent battery of tests; the most that can be expected is some degree of correlation between the two. The same psychological reputation may however be extended to the second vector if each is regarded as providing an estimate of the factor liable to error and a limited degree of correlation is considered acceptable. It is not easy to prescribe limits to the extension of such reputations.

In summary, all methods of internal factor analysis acceptable to psychologists terminate in the extraction of a number of common factors which is smaller than the number of tests, and thus divide the m-space into an f-space containing axes along which variation is relatively wide and an s-space where the dispersion is relatively small. They assume that the dimensions of interest to psychology are confined to the former sub-space and do not extend to the latter.

This assumption is open to doubt. Partial regression coefficients and discriminant functions are defined, we have seen, by their β-vectors and γ-vectors respectively as dimensions in the m-space of the original tests, but they are not necessarily related by definition to the factor space. Are there any theoretical reasons for expecting to find them contained in the *f*-space? Or is this where they are usually found experimentally? There will be opportunity for considering the theoretical question in Chapter 7, where the general principles of canonical analysis of discriminance are described (cf. especially section 2) and an experimental example will be discussed in Chapter 8 (cf. section 5).

Since the factor space is contained in the test space, factor scores can have no advantages over test scores for purposes of estimation. Some accuracy must be lost in substituting them for the test scores, even if it is not always great. Positive advantages can only be claimed for some other purpose, e.g. showing the relevance of the data to a psychological theory.

7

CANONICAL ANALYSIS OF DISCRIMINANCE

P. Slater

1. THE VARIANCE-COVARIANCE MATRICES WITHIN GROUPS AND BETWEEN GROUPS

IN Chapter 6 the N cases under discussion were treated as belonging to one group. We will now go on to consider the variation which becomes observable when they are separated into a certain number of groups, say g. 'Canonical variate analysis' as it is frequently called, or 'canonical analysis of discriminance' as I would prefer to call it, is an extension of principal component analysis to the variation observed between groups in a multidimensional space. To introduce the notation, let us start with a single variable X, used to measure N individuals forming representative samples of g groups, A, B . . . G. Any data originating from one of the groups can be indicated by one of the subscripts $a, b \ldots g$. Thus

n_g = the number of cases in group G
Σx_g = the sum of the observed measurements in these cases
$\Sigma x_g{}^2$ = the sum of the squares of the measurements

(Note that x is now used for a raw, unstandardized score.)
Let

$$\begin{aligned} n_a + n_b + \ldots n_g &= \mathrm{N} \\ \Sigma x_a + \Sigma x_b + \ldots \Sigma x_g &= \Sigma x \\ \Sigma x_a{}^2 + \Sigma x_b{}^2 + \ldots \Sigma x_g{}^2 &= \Sigma x^2 \end{aligned}$$

The totals provide a general mean $\bar{x} = \Sigma x/\mathrm{N}$, and a general total of the observed variation about it, viz.: $\mathrm{V}x = \Sigma x^2 - (\Sigma x)^2/\mathrm{N}$. We may also obtain a separate mean for each group, e.g. $\bar{x}_g = \Sigma x_g/n_g$ for group G, and find the total variation within the group as

$$vx_g = \Sigma x_g{}^2 - (\Sigma x_g)^2/n_g$$

The total observed variation within the groups may be defined as

$$vx = vx_a + vx_b + \dots vx_g$$

And an analysis of the variance of X could be carried out in these terms to test whether the group means differ from one another significantly. Vx would be broken down into the two sub-totals

vx, within groups, and
Vx–vx, between groups.

And hence the mean square variances and the variance ratio would be obtained.

Now suppose a second variable, Y is also measured in the same cases. We can obtain a general mean, $\bar{y}$, a set of group means, $\bar{y}_a, \bar{y}_b \dots \bar{y}_g$, and break down the total observed variation in Y, that is Vy, into the two sub-totals,

vy, within groups, and
Vy–vy, between groups.

The covariance of X with Y can also be treated in the same way. The sum of the products of the measurements can be obtained for each group, e.g. Σxy_g for group g; then the general sum

$$\Sigma xy = \Sigma xy_a + \Sigma xy_b + \dots \Sigma xy_g$$

The total observed covariance in X and Y is obtained as

$$\mathrm{W}xy = \Sigma xy - \Sigma x . \Sigma y/\mathrm{N}$$

The total within group g as

$$wxy_g = \Sigma xy_g - \Sigma x_g . \Sigma y_g/n_g$$

Finally Wxy is broken down into the two sub-totals $wxy = wxy_a + wxy_b + \dots wxy_g$, within groups and W$xy$–$wxy$, between groups. (Some of the terms in these expressions may be negative.)

Extending the procedure to the m variables,

$$x' = [x_1, x_2 \dots x_m]$$

we obtain firstly g sets of m means, one for each group, e.g.

$$\bar{x}'_g = [\bar{x}_{g1}, \bar{x}_{g2} \dots \bar{x}_{gm}]$$

(the terms here have two subscripts, the first denoting the group, the second the test), and next three $m \times m$ variance-covariance matrices:

S, the total, where the terms in the leading diagonal are such as Vx, and the rest such as Wxy;

T, within groups, where the terms in the leading diagonal are such as *vx*, and those elsewhere such as *wxy*; and

U, between groups = S − T

Both S and T may be used to form a matrix of correlations; and it will be convenient to distinguish between the two, as R_s and R_t.

2. PROPERTIES OF THE THREE DISPERSIONS

In Fig. 7.1 there are four partly overlapping ellipses, A, B, C and D, one centred in each of the quadrants formed by the intersection of OX and OY. Each exhibits a distribution of the same form; together they constitute a composite distribution with its centre at O. If we had simply this composite distribution before us, with no information for sub-classifying the N cases it contains, we might be satisfied with treating it as approximately uniform and elliptical, and use the variance-covariance matrix S to specify it.

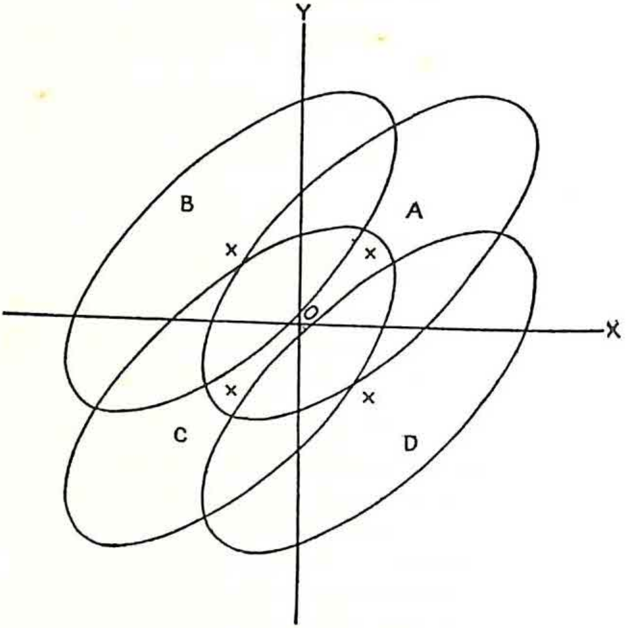


FIG. 7.1.

Then later, if information of this kind came into our possession, we might at first doubt whether distinguishing between the constituent groups is important, in view of the extent to which they overlap. Such doubts may appear more reasonable if we examine the composition of the distributions first in X and afterwards in Y, instead of attempting to view it in X and Y simultaneously: A is not distinguishable from D in X, nor B from C; conversely in Y A is not distinguishable from B nor C from D.

To investigate such possibilities systematically we must extract the matrix U. Just as the pair of measurements obtained in a single case specifies one point in the 2-space, the mean value of x and mean y for a particular group specifies one bivariate mean point. When the number of measurements increases to m the space becomes m-dimensional but the set of means for a particular group still specifies only one multivariate mean point, just as the set of measurements for an individual specifies a single point. The n_a points defined by the measurements obtained from the individuals in group A form an ellipsoidal cluster about the multivariate mean point $\bar{x}'_a$, the n_b points from group B cluster around $\bar{x}'_b$, etc. Moreover the g points specified by the group means $\bar{x}'_a, \bar{x}'_b \ldots \bar{x}'_g$ form an ellipsoidal cluster about their general mean $\bar{x}'$, which is taken for convenience as the point of origin. The form of their dispersion is specified by the variance-covariance matrix between groups, U.

Before considering the properties of U further, let us examine those of T. When the dispersion between groups has been extracted the distance between the multivariate mean points disappears. Under such conditions the four ellipses in Fig. 7.1 must become superimposed on one another with their centres all coinciding at the origin O. The residual dispersion, $T = S - U$, forms a single ellipsoid representing the dispersions within the groups collectively, provided that they all have approximately the same form. A method of testing the validity of such a postulate is given by Kendall (1957).

Returning to U, we need to note that there is an upper limit to the number of dimensions that can be occupied by the dispersion of the group means $\bar{x}'_a, \bar{x}'_b \ldots \bar{x}'_g$ about their general mean. When $m \geqslant g$ it is restricted to $g - 1$ dimensions at most, for any two points can be connected by a straight line, any three points by a plane, etc. (The general mean $\bar{x}'$ is a linear dependent junction of the group means so must be coplanar with them.) When $g > m$ the dispersion may extend into all m dimensions, but of course no further.

To distinguish between the members of any two groups as accurately as possible we need to find the axis along which the distance between the two mean points is greatest, not in absolute magnitude but *relatively to the variation within the groups.* The discriminant function already described defines this axis as the γ vector. Let us say γ_{ab} is the vector required for distinguishing between the members of groups A and B. Another vector γ_{ac} serves similarly to distinguish between the members of groups A and C, a third, γ_{bc}, to distinguish between the members of groups B and C, etc. So with g groups in all $g\,(g-1)/2$ vectors are obtained, each describing a different axis.

All these axes must lie in a space of not more than $g - 1$ or m dimensions, whichever is the lesser, and may perhaps be confined to a space of even fewer dimensions. Without noticing that we had done so, we might perhaps compute $g(g - 1)/2$ vectors which all lie approximately in a space of two dimensions and even do not diverge greatly in the second (though it would be unlikely to escape notice if they all tend to coincide along a single dimension). A finding that all the γ vectors are located in a sub-space of relatively few dimensions might be of considerable theoretical interest in certain contexts.

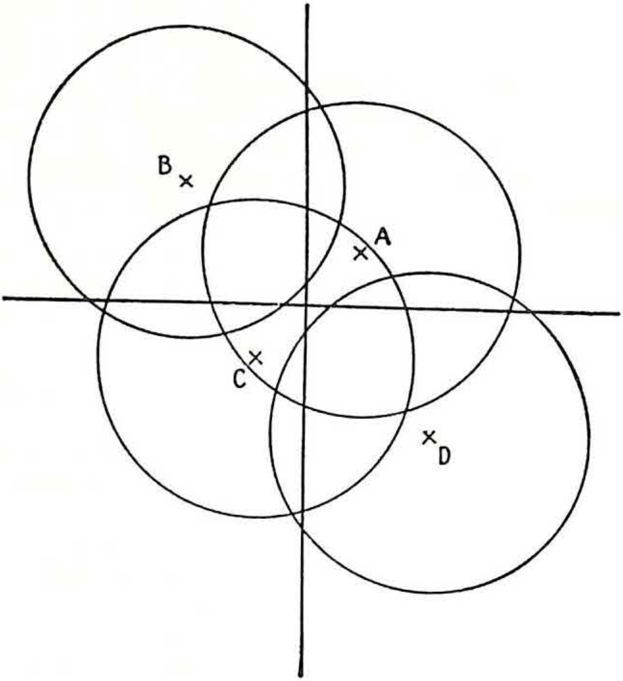


FIG. 7.2.

The mathematical procedure for ascertaining this may perhaps be understood intuitively when Figs. 7.1 and 7.2 are compared. Fig. 7.2 is obtained from Fig. 7.1 by scaling down the variation along the major axis of the dispersion within the groups, T, until it is equal to the variation along the minor axis. The ellipses are consequently reduced to circles, i.e. they come to have an equal range in all directions around their mean points. The scatter of the group means, U, is scaled down simultaneously to the same extent, and its major and minor axes become apparent. Its major axis now defines the dimension in which the variation between the groups is greatest compared with the variation within the groups.

In general an *m*-dimensional T is scaled down so that its principal axes are all of equal magnitude when it is post-multiplied, or pre-multiplied, by its reciprocal, T^{-1}. U is reduced proportionately to UT^{-1}. Its latent roots and vectors are now defined algebraically by the equation

$$[UT^{-1}]x = \lambda x$$

The first latent vector obtained by the arithmetic solution of a particular equation of this kind is the dimension where the variation between the groups is largest relatively to the variation within the groups, the second the dimension orthogonal to the first where variation is relatively next largest, etc. Let us denote these vectors as the δ vectors, the latent roots as $\delta_1, \delta_2 \ldots$ etc., and the space they define as the D-space.

Have we any reasons for expecting that the latent vectors of $U\,T^{-1}$, i.e. the δ vectors, will coincide with those of U or T or $S = T + U$, i.e. the *l* vectors obtained by principal component analysis of one of the latter matrices, or with the *f*-vectors obtained by some other method of factor analysis? There is a coincidence in Figs. 7.1 and 7.2 which is perhaps worth examining. It can be shown in general that the latent vectors of T^{-1} are the same as those of T, and its latent roots are the reciprocals of those of T. So the vector of the smallest root of T is that of the largest root of T^{-1}, etc. Now in Fig. 7.1 the dispersion of the mean points is circular, so the latent roots of UT^{-1} in Fig. 7.2 are the same as those of T in Fig. 7.1, but the relative magnitude of the roots is reversed.

For an illustration of another kind, consider the effect of including in the experiment one measurement in which there is some variation between groups, however small, and none whatever within groups. Contracting T in the manner described will eliminate all the other variables, no matter how spectacular their variances, and leave this one sufficient by itself to effect accurate discrimination.

Evidently it is possible to excogitate a number of specific instances when simpler or more complicated relationships can be demonstrated between the latent roots and vectors of UT^{-1} and those of U, T or S. But no universal mathematical relationship can be found. Mathematical considerations, that is to say, do not suggest that the ratio of the variance between groups to the variance within or the total is likely to be greater in dimensions where the latter variances are large than where they are small. Nor does any such expectation follow from the null hypothesis. There remains, however, the possibility that psychological arguments may indicate expectations in some contexts.

The possibility that the matrices obtained in a particular experiment have the same latent vectors is quite simple to check. S, the sum of U and T, and UT^{-1}, the ratio of U to T, cannot have the same latent vectors as each other, or as U or T, unless U and T have the same vectors as one another. Now the product of two symmetrical square matrices is itself symmetrical when and only when the vectors of the two are the same. So we have only to inspect the product matrix UT^{-1} and see if it is symmetrical, to discover whether U and T have the same latent vectors. We must presume that in general they have not, although admitting that there may be some instances where the vectors coincide approximately or even exactly.

3. AN EXAMPLE, ILLUSTRATING THE COMPUTATIONS

The data used here for example come from the experiment described in another chapter by R. Payne. They have been reduced to three measurements on five groups of subjects. Table 7.1 gives the composition, numbers and mean measurements of the groups, Table 7.2 the total dispersion within groups, which provides an example of T with 75

TABLE 7.1

Group	No. of Cases	Mean in Variables I	II	III
1. Normals	20	7·8513	15·6859	−4·1676
2. Dysthymics	12	14·9372	16·0880	−4·0299
3. Hysterics	8	13·4677	15·9608	−5·7374
4. Depressives	20	23·9145	16·4368	−1·9892
5. Schizophrenics	20	26·7674	20·2694	−5·8713
Sub-totals				
2 + 3 All neurotics	20	14·3494	16·0371	−4·7129
1 + 2 + 3 Neurotics and Normals	40	11·1004	15·8615	−4·4402
4 + 5 Psychotics	40	25·3410	18·3531	−3·9302
Grand Total	80	18·2206	17·1073	−4·1852

TABLE 7.2

Dispersion within groups, T

	I	II	III
I	9004·3515	−571·8267	761·3268
II	−571·8276	1853·0493	−413·4285
III	761·3268	−413·4285	954·3343

TABLE 7.3

Dispersion between groups, U

	I	II	III
I	4569·9025	842·6949	11·1115
II	842·6949	272·3667	−124·2455
III	11·1115	−124·2455	172·8772

degrees of freedom, and Table 7.3 the dispersion between groups as an example of U with 4 d.f.

A convenient way of finding the reciprocal of T is to resolve it first into the product of a triangular matrix and its transpose. Writing out the equation LL′ = T in terms of its elements as

$$\begin{bmatrix} a & 0 & 0 \\ b & c & 0 \\ d & e & f \end{bmatrix} \begin{bmatrix} a & b & d \\ 0 & c & e \\ 0 & 0 & f \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{12} & t_{22} & t_{23} \\ t_{13} & t_{23} & t_{33} \end{bmatrix}$$

we find the elements of L from

$$\begin{aligned} t_{11} &= a^2 & \therefore\ a &= \sqrt{t_{11}} \\ t_{12} &= ab & \therefore\ b &= t_{12}/a \\ t_{13} &= ad & \therefore\ d &= t_{13}/a \\ t_{22} &= b^2 + c^2 & \therefore\ c &= \text{etc.} \end{aligned}$$

Thus a simple routine is to set L under T and compute its entries column by column beginning at the top left-hand corner.

Next the reciprocal of L is found. Writing out the equation $LL^{-1} = I$ as

$$\begin{bmatrix} a & 0 & 0 \\ b & c & 0 \\ d & e & f \end{bmatrix} \begin{bmatrix} \alpha & 0 & 0 \\ \beta & \gamma & 0 \\ \delta & \varepsilon & \zeta \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

we find the elements of L^{-1} from

$$\begin{aligned} a\alpha &= 1 & \therefore\ \alpha &= 1/a \\ c\gamma &= 1 & \therefore\ \gamma &= 1/c \\ f\zeta &= 1 & \therefore\ \zeta &= 1/f \\ b\alpha + c\beta &= 0 & \therefore\ \beta &= -b\alpha/c \\ e\gamma + f\varepsilon &= 0 & \therefore\ \varepsilon &= \text{etc.} \end{aligned}$$

L^{-1} is set under L as before, but work is begun down the leading diagonal, continues down the first off-diagonal, etc. The L for the example is given in Table 7.4, L^{-1} in Table 7.5 and T^{-1}, which appears in Table 7.6, is computed as $(L^{-1})'\,L^{-1}$, i.e. L^{-1} premultiplied by its transpose.

TABLE 7.4

L, *where* LL′ = T

	I	II	III
I	94·89126	0·0	0·0
II	−6·02614	42·62317	0·0
III	8·02315	−8·56529	28·57620

TABLE 7.5

L^{-1}

	I	II	III
I	0·01053838	0·0	0·0
II	0·00148994	0·02346142	0·0
III	−0·00251220	0·00703221	0·03499416

TABLE 7.6

T^{-1}

	I	II	III
I	0·0001195885	0·0000172898	−0·0000879123
II	0·0000172898	0·0005998902	0·0002460863
III	−0·0000879123	0·0002460863	0·0012245912

TABLE 7.7

UT^{-1}

	I	II	III
I	0·56010097	0·58727150	−0·18076792
II	0·11640849	0·14738503	−0·15920743
III	−0·01601740	−0·03179883	0·18015194

Hence we can obtain UT^{-1}, which is conspicuously asymmetrical, as appears in Table 7.7. To simplify the work of obtaining its latent roots and vectors, we may substitute a symmetrical matrix for it and then apply the computing methods already mentioned in Chapter 6, Sect. 1. Putting the characteristic equation

$$[\mathrm{U}-\lambda\mathrm{T}]x=0$$

into the form $\mathrm{U}x=\lambda\mathrm{T}x$ and substituting $\mathrm{T}=\mathrm{LL}^1$ we have

$$\mathrm{U}x=\lambda\mathrm{LL}'x$$

For any non-zero λ we may premultiply both sides by L^{-1} and obtain $\mathrm{L}^{-1}\mathrm{U}x=\lambda\mathrm{L}'x$. Now if we substitute $\mathrm{L}'x=z$, so that $x=(\mathrm{L}')^{-1}z$, we find

$$\mathrm{L}^{-1}\mathrm{U}(\mathrm{L}')^{-1}z=\lambda z$$

So the matrix $\mathrm{L}^{-1}\mathrm{U}(\mathrm{L}')^{-1}$, which is symmetrical, has the same latent roots as UT^{-1}; and the latent vectors of UT^{-1}, viz. x, can be obtained from $x=(\mathrm{L}')^{-1}z$. Note that $(\mathrm{L}')^{-1}=(\mathrm{L}^{-1})'$.

The example of $\mathrm{L}^{-1}\mathrm{U}(\mathrm{L}')^{-1}$ is given in Table 7.8. Its latent roots, which are also the latent roots of UT^{-1}, are given in Table 7.9, together with its latent vectors. The latent vectors of UT^{-1} appear in Table 7.10.

TABLE 7.8

$L^{-1} U (L')^{-1}$

	I	II	III
I	0·50752173	0·28010698	−0·05443773
II	0·28010698	0·21898053	−0·11443532
III	−0·05443773	−0·11443532	0·16113579

TABLE 7.9

Latent roots and vectors of $L^{-1} U (L')^{-1}$

	λ_1	λ_2	λ_3
	0·699512	0·176284	0·011842
	l_1	l_2	l_3
I	0·691013	0·167797	−0·043209
II	0·441802	−0·127876	0·086250
III	−0·163779	0·363010	0·050356

Note: The vectors l are given in the scale $l'l = \lambda$.

TABLE 7.10

First two latent vectors of UT^{-1}, *normalized*

	δ_1	δ_2
I	0·609948	0·052309
II	0·672879	−0·035149
III	−0·418566	0·998012

Comparing the mean square variances within groups with the m.s.v.'s between groups for the three variables, given in Tables 7.2 and 7.3, we find

		for Variable		
	I	II	III	d.f.
m.s.v. within	120·06	24·71	12·72	75
m.s.v. between	1142·48	68·09	43·22	4
Variance ratio	9·516	2·756	3·397	
Probability	$<0·001$	$<0·05$	$<0·025$	

Thus there are significant differences between the groups in all three measurements. But the dispersion between groups is not thereby shown to be 3-dimensional: it could conceivably be located in two dimensions only, or even only in one. The tests of significance are inconsistent with the possibility that all the group means are clustered together too closely to be distinguishable from the general mean, but not inconsistent with the possibility that they are strung out along a straight line in the 3-space.

The canonical analysis applied to the example of UT^{-1} in Table 7.7 shows that this alternative, of collinearity, is unacceptable. The first two latent roots are both found to be significant. For a general test of the significance of the latent roots of any UT^{-1} we may enter

$$\begin{aligned}(N-g)\,\lambda_1 &\text{ as a } \chi^2 \text{ with } m + g - 2 \text{ d.f.}\\ (N-g)\,\lambda_2 &\quad .. \qquad .. \quad m + g - 4 \text{ d.f.}\end{aligned}$$

etc. Thus for the latent roots given in Table 7.9 we obtain

$$\begin{aligned}\chi_2 \text{ for } \lambda_1 &= 52{\cdot}46 \text{ with } 6 \text{ d.f., } P < 0{\cdot}001\\ \lambda_2 &= 13{\cdot}22 \qquad 4 \qquad\quad < 0{\cdot}025\\ \lambda_3 &= \ \ 0{\cdot}89 \qquad 2 \qquad\quad > 0{\cdot}75\end{aligned}$$

The dispersion between the groups is thus found to be located almost wholly on a plane in the 3-space defined by the vectors δ_1 and δ_2 and illustrated graphically in Fig. 7·3. (The co-ordinates of the group means are obtained by post-multiplying the Table 7.1 means by the Table 7.10 vectors). Using a break-down of the sum of the latent roots for descriptive purposes we might say that 78·8% of the variance between groups occurs in the dimension defined by δ_1, 19·9% in δ_2, and only 1·3% elsewhere. The main features of the dispersion are that the normals and the two neurotic groups, hysterics and dysthymics, are centred fairly closely together and that the two psychotic groups, the depressives and the schizophrenics, are well separated both from the non-psychotic groups and from one another.

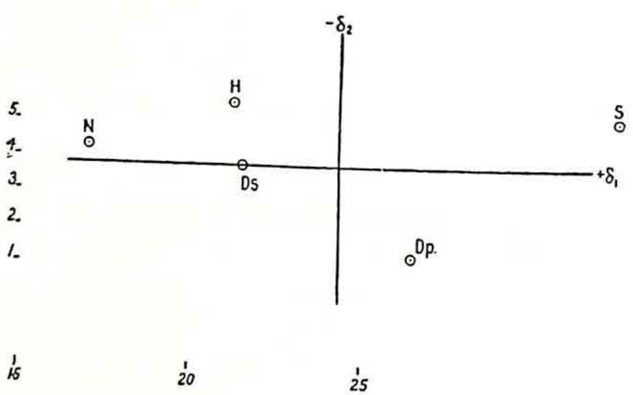


FIG. 7.3. Principal Components of the Dispersion between Groups.

Group means:

Group	δ_1	δ_2
Dp. Depressives	26·48	$-1{\cdot}31$
Ds. Dysthymics	21·62	$-3{\cdot}81$
H. Hysterics	21·36	$-5{\cdot}58$
N. Normals	17·09	$-4{\cdot}30$
S. Schizophrenics	32·42	$-5{\cdot}17$
Axes pass through the general mean	24·38	$-3{\cdot}82$

4. OTHER FEATURES OF THE EXAMPLE

Much the same conclusions can be drawn from an examination of the discriminant functions. To compare firstly the two neurotics groups we set down the mean differences obtained from Table 7.1 and premultiply by T^{-1} (Table 7·6) to arrive at the discriminant function, viz.:

	Mean differences (d)	Discriminant function (γ)
I	1·4695	0·00002782
II	0·1272	0·00052191
III	1·7075	0·00199310

This computation gives the elements of γ on a scale suitable for the next step: $d'\gamma = D$, which leads to $D = 0{\cdot}00351050$ here. The test of significance then takes the form

Source of variance	*Sum of sqrs.*	*d.f.*
Difference between group means	$\left\{\dfrac{n_a n_b}{n_a + n_b}\right\} D^2$	m
Variance within groups	D	$N-m-g$

(Where n_a and n_b are the numbers in the groups compared). And the mean square variances are found to be 0·00001972 between groups and 0·00004876 within.

Concluding that the variables under consideration do not discriminate between these two groups we may combine them together and proceed to the next comparison, of the neurotics with the normals. Means for all neurotics are listed in Table 7.1. Proceeding as before we obtain $D = 0{\cdot}00609547$ for this comparison, m.s.v.'s = 0·00012385 and 0·00008466 respectively, and thus again find no significant difference.

We should not be mistaken, now, in concluding that there are only three well-established points to consider in the dispersion between groups: the mean points for the endogenous depressives and the schizophrenics and a combined mean for normals and neurotics together. There is thus no difficulty in understanding why the dispersion between groups is located on a plane.

In looking for axes to which to refer the variation between these means we are not restricted to the use of δ_1, and δ_2. We may perhaps allow psychological or administrative considerations if they are sufficiently cogent to persuade us of the advantages of using another pair. Using a discriminant function again to compare the schizophrenics with the depressives we have

d	and hence	γ_1
−2·8529		−0·00074872
−3·8326		−0·00139313
3·8821		0·00406164

$\mathrm{D} = 0{\cdot}02324305$ and the m.s.v.'s $= 0{\cdot}00180080$ and $0{\cdot}00032282$ respectively. The variance ratio is $5{\cdot}578$, and $\mathrm{P} < 0{\cdot}01$. And comparing the psychotics with the others

d	and hence	γ_2
$-14{\cdot}2406$		$0{\cdot}00170126$
$-2{\cdot}4916$		$0{\cdot}00186641$
$-0{\cdot}5100$		$-0{\cdot}00001423$

Here the variance ratio is $13{\cdot}858$ and $\mathrm{P} < 0{\cdot}001$. When these two vectors are normalized they become

	γ_1	γ_2
I	$-0{\cdot}1718$	$0{\cdot}6736$
II	$-0{\cdot}3196$	$0{\cdot}7390$
III	$0{\cdot}9318$	$-0{\cdot}0056$

The angle they subtend is approximately 70° ($\gamma'_1 \gamma_2 = \cos\theta = 0{\cdot}3572$, $\theta = 69^\circ\ 10'$). They are the axes preferred by Payne.

A few supplementary details need perhaps to be added to the information about the variables given by Payne. The original observations from which they are derived are scores on a battery of 54 tests. The sums of squares and products of deviations from the general means for all $\mathrm{N} = 80$ cases were scaled down to a 54×54 matrix of correlations, R_s. A centroid factor analysis was applied, with estimated communalities substituted for the units in the leading diagonal of R_s, and three factors extracted. The unrotated loadings form a matrix, f, with 54 rows and 3 columns.

The three variables are the factor scores for these three factors. For the purposes of the analysis it is unnecessary to compute them for each case individually; we only need the data in Tables 7.1 to 7.3, relating to group means and dispersions. Using the definitions already given

x_i for an individual's score in test X
$\bar{x}$ for the general mean in all N cases
$\mathrm{V}x = \Sigma(x_i - \bar{x})^2$, the sum of squares of deviations from $\bar{x}$,

let us write

$$s_i = x_i/\sqrt{\mathrm{V}x/(\mathrm{N}-1)}$$

and

$$\bar{s} = \bar{x}/\sqrt{\mathrm{V}x/(\mathrm{N}-1)}$$

so that the individual's standardized score in X can be expressed as $s_i - \bar{s}$, i.e. as a difference between two terms, and the row vector of his standardized scores in the 54 tests as $s'_i - \bar{s}'$, i.e. as a difference between two vectors, one referring to him personally and one to the general mean. Analogously

$$\bar{s}_g = \bar{x}_g/\sqrt{\mathrm{V}x/(\mathrm{N}-1)}$$

gives the characteristic element in the row vector $\bar{s}'_g$ of means for group G.

The three factor scores for an individual could now be found, if required, from $s'_i f$; but the five rows of group means in Table 7.1 can also be obtained directly as $\bar{s}'_g f$, and the final row of means for all 80 cases as $\bar{s}' f$. The total dispersion of the factor scores between groups, which is given in Table 7.3, can be computed as

$$\mathrm{U} = \Sigma n_g f' \bar{s}_g . \bar{s}'_g f - \mathrm{N} f' \bar{s} . \bar{s}' f$$

That is to say, each row of Table 7.1 is premultiplied by its transpose to form a 3 × 3 matrix, the elements of which are multiplied by the number of cases in the group. The 5 matrices so obtained are summed. Lastly a similar matrix formed from $\bar{s}' f$ is subtracted. The total dispersion within groups given in Table 7.2 is obtained as the difference

$$\mathrm{T} = \mathrm{S} - \mathrm{U}$$

after S, the total dispersion about the multivariate mean for all groups combined has been found as

$$\mathrm{S} = (\mathrm{N} - 1) f' \mathrm{R}_s f.$$

5. RATIONALE

A case for employing the analysis described in the two foregoing sections can be presented without postulating that the f-vectors (or any others derivable from them by rotation) have any particular psychological connotation.

Since there are only 80 cases to allocate and 54 scores are recorded for each, it is almost a foregone conclusion that a set of linear combinations of the scores can be found to provide an accurate allocation procedure. The immense labour of the computations will lead to a trivial result, for however satisfactory it appears to be it will not be safe to generalize. But to show that a greatly reduced number of measurements will serve the same purpose may be of interest; so some preliminary procedure is worth considering for reducing the scores to an appropriately small number of variables.

The centroid factor analysis serves this purpose as a rough approximation to a principal component analysis, cf. Burt (1940). The three columns of f, treated as vectors, define approximately the first three principal axes of R_s—the three orthogonal dimensions of variation where differences between individuals are most conspicuous and confirmed most reliably by the cumulative evidence of all the tests. If the search for evidence of discrimination is to be restricted to only three out of the 54 dimensions where it may be found, these have a claim to priority. The claim would have been better of course if the principal components had been located exactly; and the subsequent computations would have been simplified as exact orthogonality would have been obtained; but as a matter of historical fact a centroid analysis was used.

The canonical analysis demonstrates that the group means form a two-dimensional scatter as shown in Fig. 7.5. Purely from considerations of administrative expediency we might prefer to substitute two other axes for the principal axes of this dispersion, and employ firstly a measure for distinguishing psychotics from non-psychotics, viz. γ_1, and secondly when applicable a measure for subclassifying the psychotics into schizophrenics and depressives, γ_2. The successive operations involved in obtaining the two measures for an individual are defined algebraically as

$$(s'_i - \bar{s}')f\gamma$$

Consequently the product matrix $f\gamma$ (cf. Payne's chapter) defines the two required vectors in the original m-space.

In proceeding further to consider the relevance of the evidence to the psychological theories proposed by Payne several points seem worth noting. The tests were originally selected to ensure that their cumulative evidence would provide reliable measures of two psychological traits, one believed to distinguish psychotics from non-psychotics and the other schizophrenics from depressives. A search restricted to the first three axes of R_s therefore seems sufficient for the purpose of investigating the theory. Support for it comes from the findings that the constellation of the group means is restricted to two dimensions within the three investigated, and separates the groups in the expected manner—without, for instance, discriminating between neurotics and normals. Further support is obtainable from a detailed examination of the elements in $f\gamma$, which tends to show that the weights assigned to the tests for estimating individual differences in the discriminant dimensions accord with their postulated relevance to the two psychological traits.

However, it should not be forgotten that the search for discriminance has been very incomplete. Only three out of 54 dimensions have been explored. It would thus be well to avoid drawing negative or restrictive conclusions, such as that the tests serve to measure no other traits than the two described, or that they provide no way of distinguishing hysterics from dysthymics. We cannot even be sure that better discriminants for the proposed purposes do not occur in some other sub-space which has not been investigated. The vectors found are lines of best fit only within the chosen sub-space; they are optima in a very restricted sense.

The danger that valuable evidence may be overlooked as a result of an incomplete factor analysis will be illustrated in Chapter 8.

8

A RE-EXAMINATION OF SOME DATA COLLECTED BY H. P. HILDEBRAND

P. Slater

1. HILDEBRAND'S EXPERIMENT

THE data now to be examined were collected by H. P. Hildebrand during a course of studies for a Ph.D. at this Institute. His findings, based mainly on a centroid factor analysis, were published first in his thesis (1953) and later in an article in the British Journal of Psychology (1958). Eysenck has quoted them extensively (e.g. 1957) and asked me to make some further analyses of parts of the data, using the methods described in Part II. Most of this was done during 1956–7; the electronic computation by the National Physical Laboratory under the supervision of Mr. J. G. Hayes. L. H. Storms, who was pursuing a course in statistics here at that time, used other parts of the data not too bulky to be treated by electric calculator for a parallel analysis, and published his findings in the *J. Ment. Sci.* in 1958. Anyone who goes to the trouble of reading all three texts is likely to find differences in the authors' interpretations of the results and may even come to the opinion that the different methods of analysis conflict. In this account I shall attempt to show where the differences originate and suggest a synthesis.

Hildebrand's research exemplifies the hypothetico-deductive method. To understand what he did one must know what postulates he had in mind and how he proposed to test them. He supposes that introversion-extraversion (hereafter denoted I–E for short) is a major dimension of personality, and follows Jung (1924) in considering that it occurs in its entire range among normal subjects. He does not accept Freud's opinion (1920) that neuroticism is associated with introversion and normality with extraversion. He adopts Eysenck's view that the distinction between neurotics and normals is to be found in a dimension of neuroticism which is orthogonal to I–E, and follows him further in supposing that I–E serves to subclassify neurotics though it does not distinguish them collectively from normals. The anxiety states are expected to occur preponderantly among introverts and the hysterics among extraverts.

His experimental groups consisted of

No. of cases	Diagnosis
25	Anxiety states
10	Obsessional neuroses
10	Reactive depressions
25	Conversion hysterics
20	Psychopathic states
55	Mixed neuroses
25	Normals
170	Total

The neurotic subjects were in-patients at Belmont Hospital, Sutton, Surrey. They were all males, aged between eighteen and fifty, with IQ's above 90, showing no signs of organic lesion or psychotic illness, and not undergoing physical treatment by ECT or insulin therapy. The 'mixed neuroses' were cases where the diagnosis was not clear-cut. The control group of twenty-five normals (the H-controls) was drawn from soldiers passing through an Army Reallocation Centre in the neighbourhood. This selection may have been unfortunate. Reallocation is provided by the Army for men found unsuitable for the duties to which they were posted as new recruits, and the centres are places where all sorts of misfits collect. Eysenck replaced this group later by a sample of twenty-two men (the E-controls) mostly working-class people such as porters and lorry drivers, chosen as being within the same age range as the neurotics, who were contacted individually and invited to attend the Institute for testing—none refused.

The original test programme was to have yielded thirty-two measurements in each case, but some omissions and some amalgamations of tests or test scores reduced the number to twenty-two. They are all described by Hildebrand as *objective*, that is to say, they were scored by a standard set of instructions not requiring subjective interpretation by the experimenter. Self-rating scales are thus included. There were three measures of intelligence: Progressive Matrices, the Mill Hill Vocabulary and the Nufferno Power test; four of neuroticism: the Maudsley Medical Questionnaire and the lie scale, an Annoyances test based on the Pressey X–O scale, and the combined D and C scales from the Guilford-Martin inventory; the remaining fifteen were expected to exhibit variation in I–E. They are enumerated in full in Table 8.1.

TABLE 8.1

Data extracted from Hildebrand (1953 *and* 1958)

Test	Variance Ratio	Factor loadings	
		N	I–E
A. *Also administered to the second control group*			
Mill Hill Vocabulary	1·765	−0·039	0·110
Maudsley Medical Questionnaire	1·154	0·739	−0·257
Lie scale	2·665*	−0·380	−0·371
S	1·448	0·739	−0·472
T scales from the Guilford-Martin	1·776	0·467	0·099
D Inventory (D and C scales totalled	1·185	0·831	0·057
C in the factor analysis)	1·719		
R	3·944**	−0·246	0·563
Annoyances	2·131	0·429	−0·012
Weighted wrong Directions and	1·642	0·257	0·360
Crossed Lines from Porteus Maze Test	1·800	0·067	0·425
B. *Not administered to the E-controls*			
Progressive Matrices	1·084	0·179	−0·014
Nufferno Power test	1·080	0·098	0·012
Sex Humour	0·916	0·301	0·222
Static Ataxia, Total distance	1·153	0·086	0·180
,, ,, Reversals	1·011	−0·041	0·073
Suggestibility, Total distance	2·414*	0·176	0·052
,, Reversals	3·936**	0·021	0·211
Body build Index	0·570	0·015	−0·255
Starting time	2·251	−0·167	−0·195
Total time from the Porteus	1·246	−0·062	−0·071
Lifted pencils Maze test	1·029	0·070	0·006
Leg persistence	2·791*	0·113	−0·066
C. *Not included in the factor analysis*			
Extrapunitive humour	1·108		
Incongruous humour	0·953		
Food aversion	0·251		
Aesthetic differences	1·266		
Static Ataxia, Forward distance	1·945		
,, ,, Backward ,,	1·039		
Suggestibility, Forward ,,	2·007		
,, Backward ,,	(n.s.)		
Wavy Lines (Mazes)	1·570		

Note: The variance ratio tests the significance of the differences between the means of the neurotic groups. *0·05 < P < 0·01 **P < 0·01

Hildebrand's argument leads him to expect that the dimensions of I–E and neuroticism can be located by a factor analysis restricted to the data remaining after the anxiety states, the hysterics and the normals have been excluded to provide criterion groups. Then when factor scores for I–E and neuroticism are calculated for everyone it should be found that the means for anxiety states and hysterias differ significantly on the I–E scale, where the normals are to be found in a central position with the mixed neurotics; while on the neuroticism scale the normals should be separated from the other groups.

The correlations computed from the restricted data were subjected to a centroid analysis to obtain first estimates of the factor loadings for three factors. Two iterations were carried out by Lawley's maximum likelihood method (1950); and the resulting residuals were tested for significance by Bartlett's test (1950), and found to be still significant. A fourth factor was accordingly extracted. On the basis of this evidence Hildebrand concludes that all four factors are significant. Two rotations, he adds, were sufficient to adjust the factors to psychologically meaningful positions. The resultant loadings for the first and third factors are reproduced in Table 8.1.

The four factors together account for 32% of the total variance. The first accounted for 12% and was, to quote, 'easily identified as a neuroticism factor. The four variables with the highest loadings differentiated significantly between normals and neurotics. Moreover, previous studies had shown these tests to be highly saturated with neuroticism, emotional instability or similar factors. The second accounted for 9%. It was confidently identified as an intelligence factor. The first three variables were the three tests of intelligence, and there are high correlations with mistakes and length of time on the Porteus Maze test. The third accounted for 6%. Critical ratios were calculated for the scores of the most highly loaded of the variables between the hysterics and the dysthymics. The tests do in fact discriminate between the groups.' (The tests cited are the Guilford S and R scales, Suggestibility reversals, and the WWD and X lines scores on Porteus.) 'For these reasons the factor was identified as an I–E factor. The fourth factor extracted was identified as arising from the intercorrelations of the static ataxia and suggestibility scores. It does not seem possible to identify it more meaningfully.'

Multiple regression equations were then computed for estimating the I–E and neuroticism scores, and the scores were calculated for every individual.[1] The group means were found to be

Group	Neuroticism	I–E
Anxiety states	5·949	1·407
Obsessional neuroses	5·424	1·082
Reactive depressions	4·647	1·295
Conversion hysterias	4·866	2·241
Psychopathic states	5·916	2·637
Mixed neuroses	5·364	1·593
H-control normals	3·402	2·875

[1] This part of the work, which is exceedingly laborious and exacting, was done by

Hildebrand claims that apart from the position of the normal group at the extravert end of the I–E continuum, 'inspection of the means demonstrates clearly that the distribution follows the prediction'. He points out that the representativeness of his control group is open to question, and remarks, 'It is possible that if a less homogeneous and more representative civilian sample had been tested the predictions would have been fulfilled.'

He is, however, quite reserved in his appraisal of the importance of the I–E factor. 'The experiment was purely on a descriptive basis,' he comments. 'Whether or not the factors found, particularly the I–E factor, may be considered to have hypothetical causal instances[1] underlying them is not a question that can be answered on the basis of the data here presented. A certain amount of caution must be used before the factor can be used as a basic postulate in personality research. Although the experiment was designed to amplify the factor as much as possible, yet the total variances it accounted for was only 6%, only 1% more than that of a factor found impossible to identify satisfactorily.'

2. EYSENCK'S USE OF THE FINDINGS, AND HIS REASONS FOR TAKING A SECOND CONTROL GROUP

Eysenck (1947) goes much further. He describes extraversion-introversion as a type concept with the 'highest level of generality'. He connects the concept with Hull's law of reactive inhibition (1943) in a formulation linking personality theory with learning theory: 'Individuals in whom reactive inhibition is generated quickly and strongly, and dissipated slowly, are thereby predisposed to develop extraverted patterns of behaviour, and hysterico-psychopathic disorders in cases of neurotic breakdown; conversely individuals in whom it is generated slowly and weakly, and dissipated quickly, are thereby predisposed to develop introverted patterns of behaviour, and dysthymic disorders in cases of breakdown.' He follows this up with a great array of evidence, which needs not be reviewed here.

To confirm the theory evidence would seem required showing

(1) That a certain degree of I–E can be inferred in an individual from the consistency of his results in several relevant tests.

(2) That a certain degree of liability to reactive inhibition can likewise be inferred from the consistency of his behaviour in various learning situations.

Miss le Prevost. Only the eight highest regression coefficients were used in computing the factor scores.

[1] The phrase is borrowed from Eysenck 1953a.

(3) That assessments of the two characteristics will be found related in a specified way.
(4) That among neurotics both assessments will be found related to diagnostic differences in a specified way.

Compared with other sources, particularly the psychoanalytic ones, which might be cited in support of the view that many aspects of behaviour reveal differences in I–E, Hildebrand's experiment has the advantage of specifying a weighted summation of scores on a particular set of 'objective' tests as appropriate for measuring it. Moreover, Hildebrand adduces evidence that these measurements distinguish between neurotics in the specified way. It is not surprising that Eysenck relies on Hildebrand's findings as his main support for the part of the theory concerning I–E as a personality trait.

The greatest weakness of the findings from Eysenck's point of view is that they do not accord with the expectation about the mean score of normal people on the I–E scale. The H-controls' mean places them at the extraverted end, beyond even the psychopaths, while the remaining groups lie fairly close together near the other end, and the difference between the anxiety states and the hysterics is not conspicuous. The variance between the groups could be broken down as follows

Source	Sums of squares	d.f.
Between Normals and Neurotics	26·47	1
Between Psychopaths and other neurotics	17·81	1
Among other neurotic groups	14·77	4
Variance within groups, total	270·57	163

And the variance among the neurotic groups excluding the psychopaths might accordingly be dismissed as non-significant (the F ratio being 2·22). However, when the anxiety states are compared directly with the hysterics a significant difference is found on a one-tailed test in the direction required by the hypothesis, and this may be accepted as sufficient confirmation.

Three explanations might be put forward to account for the anomalous mean score of the H-controls: the least disturbing is that the H-controls are not a representative sample of normal men. The alternatives are that the scale is not correctly identified as an I–E factor, or finally, if it is, that Jung's view of the frequency distribution of I–E in the normal population must be discarded in favour of Freud's. If a mean nearer the general average for all neurotics is obtained when another sample of controls is drawn from a less suspect source, Eysenck argued, the new evidence will decide in favour of the first alternative and the original theory will be confirmed. This was his reason for taking the second sample, the E-controls, and giving them only the reduced

TABLE 8.2

The matrix of correlations between the tests given to the neurotic groups and the E-*controls calculated from sums of squares and products of deviations from group means, i.e.* R_t

Test	2	3	4	5	6	7	8	9	10
1. MHV	−0·1547	−0·0127	−0·0942	0·0721	0·1489	−0·1261	−0·1009	−0·1601	−0·0604
2. MMQ		−0·2395	0·4085	0·4867	0·2839	0·6278	−0·1966	0·0839	0·0582
3. Lie scale			−0·1730	−0·1673	−0·2662	−0·2241	0·0488	−0·0220	−0·1933
4. Annoyances				0·2652	0·3522	0·4106	−0·0524	0·0931	−0·0355
5. S					0·3667	0·5342	−0·4195	0·0767	−0·0118
6. T						0·5360	−0·2153	0·0675	0·1369
7. D & C							−0·0846	0·1646	0·1392
8. R								0·1925	0·1214
9. WWD									0·4156
10. X lines									

TABLE 8.3

Group means, etc., in tests given to the E-*controls*

Test	Anx.	Obs.	Dep.	Hys.	Psychop.	Mx.	E-controls	s.d.	F
MHV	19·72	24·00	20·10	18·72	22·10	19·95	19·59	5·786	1·438
MMQ	24·00	20·80	20·70	21·16	22·95	20·47	8·68	6·708	12·770**
Lie scale	5·40	5·70	7·30	6·56	4·00	5·82	5·82	2·784	2·260
Annoyances	16·96	20·10	16·20	18·56	21·70	17·33	16·82	6·155	1·939
S	33·08	30·90	26·70	24·48	27·10	29·44	24·09	10·670	2·220*
T	36·64	43·20	33·40	34·92	41·85	38·40	31·41	11·263	2·483*
D + C	79·28	76·80	62·50	65·08	81·45	74·71	40·36	25·403	6·954**
R	27·00	28·60	27·10	36·60	39·20	33·24	33·82	11·667	3·125**
WWD	16·40	13·90	11·00	17·76	20·70	15·73	15·18	9·584	1·496
X Lines	9·00	6·30	10·40	12·80	11·10	8·78	10·09	7·114	1·514
Factor score:									
Neuroticism	1·950	1·830	1·433	1·495	1·943	1·732	0·960	0·7548	4·739**
I–E	0·0411	0·1903	−0·0969	0·3262	0·7552	0·2389	0·3962	0·7338	2·441*

Notes: The s.d. tabulated is the general estimate of the standard deviation within groups, viz. $(vx/(N-g))^{\frac{1}{2}}$

F is tabulated as the ratio (m.s.v. between groups)/(m.s.v. within groups). The control group is included.

battery of tests (cf. Table 8·1) which would be sufficient, according to Hildebrand's factor analysis, to provide reliable assessments of neuroticism and I–E.

Factor scores calculated from the reduced battery are of course not exactly comparable with those from the complete battery, quoted above. To obtain comparable estimates for all the groups the procedure appropriate for the E-controls must be extended to the others. Table 8.2, which gives the correlations between the tests, is completed by symmetry with units in the leading diagonal to form a matrix R. Neuroticism and I–E are treated as criteria, as explained in section 2, Chapter 6; their correlations with the tests (y) are provided by the factor loadings in Table 8.1. The values of x which satisfy the equation

$$x = \mathrm{R}^{-1}y$$

provide the partial regression coefficients, $x = \beta$, for estimating neuroticism and I–E, as defined by Hildebrand's vectors, from the test scores on the reduced battery. They are found to be

Test	for Neuroticism	for I–E
MHV	0·0408	0·1400
MMQ	0·1511	−0·1894
Lie scale	−0·1166	−0·2510
Annoyances	0·0108	0·0156
S	0·1906	−0·3355
T	−0·0271	0·0263
D + C	0·2985	0·2493
R	−0·0295	0·1842
WWD	0·1238	0·1597
X lines	−0·0701	0·1062
Multiple correlation	0·7548	0·7338

The multiple correlations $(y'\beta)^{\frac{1}{2}}$, show that both estimates are reasonably reliable. The same statistics give the s.d.'s of the estimates within groups, since

$$y'\beta = \beta'\mathrm{R}\beta$$

Factor scores do not have to be calculated for each individual to find the means for groups. The procedure described in section 4, Chapter 7, can be followed. The mean test scores for the groups are given in Table 8.3. Each column of entries constitutes a vector $\bar{x}_g$; division by the corresponding entries in the column of s.d.'s gives $\bar{s}_g$; and the product $\bar{s}'_g\beta$ gives the mean factor score, listed at the foot of the table.

The bivariate distribution of the means is shown graphically in Fig. 8.1. The axes are set through the general mean point for all neurotics, which coincides almost exactly with the mean point for the mixed neurotics. The separation of the anxiety states from the hysterias along the I–E axis, though not conspicuous, is in the right direction, the deviation of the anxiety states towards the introverted end of the scale being more marked than that of the hysterias towards the extraverted end.

When the corresponding graph given by Hildebrand is compared with it, the alteration in the location of the control group is the main difference found. There, as already mentioned, the mean for the H-controls lies at the extraverted end of the I–E scale, beyond the mean of the psychopaths. Here the mean for the E-controls, while still well away from the means of the neurotics on the neuroticism scale, lies not far from their general mean on the I–E scale. One cannot accept both groups as representative of normal men, and doubt falls on the H-controls in view of their suspect origin. The E-controls conform much more closely with the expectations deduced from the general theory than the H-controls. Therefore confirmation for the theory has been obtained.

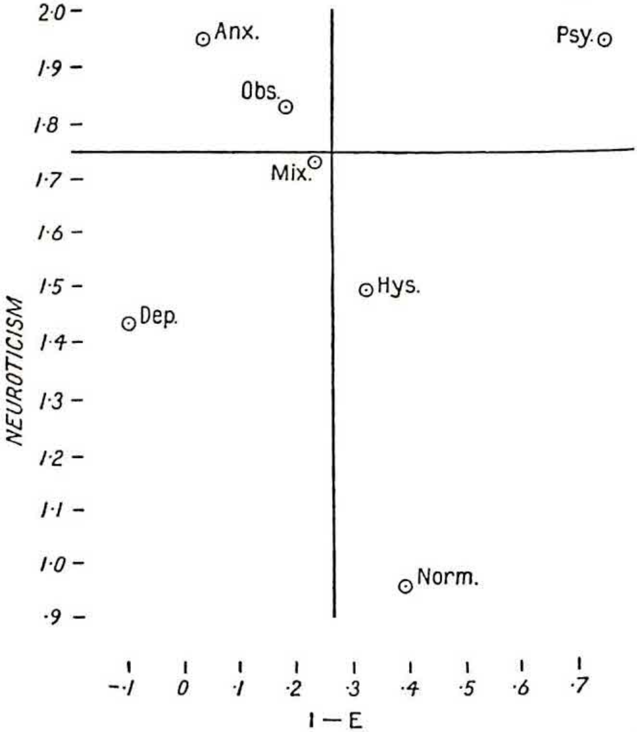


FIG. 8.1. The Distribution of Group Means for Neuroticism and I–E estimated from Scores on the Reduced Battery of Tests. The Axes are drawn through the Bivariate Mean Point for all Neurotic Cases (N = 1·744, I–E = 2646).

In my opinion Hildebrand's investigation and Eysenck's continuation of it combine to form a very competent as well as important research of the factor-analytic kind. It is not open to criticism except in terms extending to methods of centroid analysis in general. Their use of the method compares well with other investigators'.

As a result a weighted summation of scores on a battery of tests (i) is accepted as measuring I–E, (ii) is found to discriminate between neurotic groups. It is accordingly approved for use in further researches. If they show (iii) that it is also associated with liability to reactive inhibition in the postulated way the outcome of the larger project to which they belong will be judged satisfactory on the whole. A connection indirectly linking differences in reactive inhibition with diagnostic differences will have been formed and explained. The argument will not be invalidated even if evidence is adduced, now or later, to show that some other weighted summation of the same or different tests either provides a better measure of I–E or discriminates more efficiently between neurotic groups.

3. THE DISCRIMINANCE OF THE TESTS

Hildebrand has not convinced me that the vector he has chosen for measuring I–E is the optimum one for this purpose. But this comment is not a criticism of his treatment of the data. I know of no line of reasoning which leads to an indisputable optimum for the vector defining a postulated factor such as I–E. I am therefore likewise debarred from proposing any other vector as preferable. Eysenck's acceptance of Hildebrand's findings has established a consensus of opinion I can see no advantage in disturbing. On the contrary there is an advantage in having a scale for measuring I–E precisely even if arbitrarily defined; its properties can then be examined in detail.

When we enquire what test, or what linear combination of the tests is best for discriminating between the neurotic groups, we set out to search the data for evidence of quite a different kind. Although Hildebrand anticipated that his I–E factor, when identified, would be found to distinguish between the neurotic groups in a particular way, he did not postulate, as far as I am aware, that it would provide the best weighted summation of the tests for this purpose. So there is no cause for disappointment in finding that the merits of the I–E scale are neither negligible nor outstanding in this respect.

In other words, we are now propounding a question he never considered, when we enquire what would be the effects of substituting his f-vectors for the γ- or δ-vectors which would give the best results. All three define sets of weights for use in summation. When optimum weights are used, a summation of two measurements must give a better estimate than either alone, and each additional measurement will increase the accuracy of the estimate even if only to a negligible extent. But when unsuitable weights are used adding poorer measurements to better ones may produce a less accurate estimate than omitting them. The f-loadings tend to have this effect.

For discriminating the E-controls from the neurotics collectively the best single test in the ten-test battery is the Maudsley Medical Questionnaire. The second is the D plus C scale from the Guilford-Martin inventory. The neuroticism index, although it includes the MMQ and D plus C scales, comes third. The variance between controls and neurotics, expressed as a proportion of the total sum of squares, is

29·8% on the MMQ
16·7% ,, ,, D + C scale
11·4% ,, ,, Neuroticism index

On the other hand the discriminant function is significantly better than the MMQ alone; the proportion of its total sum of squares due to the variance between controls and neurotics is 38·1%. Its vector is given in Table 8.4.

The I–E index is more successful; in discriminating between the neurotic groups it comes a close second to the Guilford R scale. The variance between groups, expressed as a proportion of the sum of squares for all neurotics, is

11·8% on the R scale
10·1% ,, ,, I–E index

If we take for comparison the discriminant function maximizing the difference between the anxiety states and the hysterias, we find the percentage only rises to 16·0; the increment over the discrimination provided by the R scale is not significant. Group means measured along this axis are

Anxiety states	−0·7337
Obsessions	0·3832
Depressions	0·0728
Hysterias	0·6596
Psychopathics	0·1777
Mixed neuroses	0·0255
E-controls	0·5478

(When the origin is taken at the mean for all neurotics and the s.d. within groups is equated to 1.) It is interesting to see that an axis can be found which places the anxiety states and the hysterias at opposite poles; and also that in this dimension the mean for the normal cases (the E-controls) lies well away from the mean for all neurotics and close to the mean for the hysterias.

In the complete set of discriminant functions obtained by pairing each group with every other the normal group is found significantly different from all the rest. But the vectors for every comparison do not seem of sufficient general interest to be worth tabulating separately. The only vector relating to the normal group given in Table 8.4 is the one already mentioned, which discriminates them from neurotics collectively.

The remaining results are summarized below:

	Obs.	Dep.	Hys.	Psy.	Mix.
Anxiety states	—	—	+	++	—
Obsessions		—	—	—	—
Depressions			—	++	—
Hysterias				+	—
Psychopathics					+
Mixed neuroses					

Note: +, discrimination significant at the 5% level
++, ,, ,, ,, ,, 1% ,,

Evidently the psychopaths form an isolated group of cases. They are distinguishable even from the hysterics. The obsessionals are the only cases not clearly distinguished from them; and this is not so much because of any close resemblance (cf. the constellation of group means in Figs. 8.2, 8.3 and 8.4) as because of the shortage of obsessional cases. Apart from the psychopaths, there is very little evidence that the remaining groups can be distinguished from one another. The only significant discriminant is the one already discussed, which distinguishes anxiety states from hysterias. There is of course no justification for any categorical conclusion that any of these groups are indistinguishable. A theoretical possibility always remains open that some other tests discriminate, and in section 6 below it is shown that some of the tests omitted from the reduced battery actually do so.

All the vectors giving significant discrimination between one group of neurotics and another are listed in Table 8.4

4. THE CANONICAL AXES OF THE DISPERSION IN THE D-SPACE

Since the normals and the psychopaths are the only isolated groups and the remaining groups tend to cluster together, we might be led to expect that the D-space between the group means will not be more than two-dimensional. The expectation is strengthened because the tests were originally selected by Hildebrand to provide discrimination in only two dimensions, neuroticism and I–E, and were screened again on the evidence of his factor analysis and only retained if their loadings indicated that they acted as measures of either or both. The rather surprising finding is that there are three significant independent dimensions of variance between the group means.

The latent roots and vectors of the dispersion are given in the first part of Table 8.5. The tests of significance show that for

δ_1, $\chi^2 = 103{\cdot}32$ with 15 d.f.
δ_2 ,, $= 39{\cdot}93$,, 13 ,,
δ_3 ,, $= 28{\cdot}64$,, 11 ,,

TABLE 8.4

Discriminant functions from the ten-test battery.

Test	Vectors of the discriminants between					
	Normals and	................Psychopaths and			............Mixed	Anxiety states
	Neurotics	Anxieties	Depressives	Hysterics	Neurotics	and Hysterics
1. Mill Hill Vocabulary	0·3511	−0·3559	−0·3852	−0·4512	0·4037	0·0953
2. Maudsley Med. Quest.	1·4943	0·0932	0·0721	0·1372	0·2600	−0·0440
3. Lie scale	0·2562	0·2638	0·7205	0·5832	−0·3413	−0·3194
4. Annoyances	−0·4823	−0·5424	−0·3208	−0·1177	0·4087	−0·4247
5. S	−0·5709	0·3580	0·1767	0·1624	−0·3715	0·1956
6. T } scales from the Guilford-Martin	0·1329	−0·2808	−0·2119	−0·0554	0·0005	−0·2254
7. D + C } Inventory	0·4425	0·0672	−0·2549	−0·4444	−0·0242	0·5115
8. R	0·0705	−0·5794	−0·6433	−0·1628	0·2319	−0·4166
9. WWD } Porteus	0·0919	−0·1999	−0·6822	−0·3461	0·3075	0·1462
10. X lines } Mazes	−0·1945	0·0071	0·4672	0·4576	0·0044	−0·4505
	Variance ratios from the tests of significance					
	5·718	2·883	2·904	2·259	2·451	2·275

Note: To simplify reading the γ-vectors have been given here multiplied by 100.

TABLE 8.5

Latent roots and vectors of the dispersion of the means in the D-*space*

PART 1.	δ_1	δ_2	δ_3
Latent roots	0·64577	0·24954	0·17903
Variable	Latent vectors, normalized		
1. MHV	0·1759	0·3450	−0·1031
2. MMQ	0·6616	−0·1960	0·4354
3. Lie scale	0·0715	−0·4277	0·3086
4. Annoyances	−0·2081	0·4318	0·2592
5. S	−0·2544	−0·2063	−0·2585
6. T	0·0414	0·2101	0·0003
7. D + C	0·2131	0·0594	−0·4710
8. R	0·0212	0·4663	0·2096
9. WWD	0·0710	0·3502	−0·1219
10. X lines	−0·1064	−0·2071	0·5384
PART 2.			
Group means			
Group			
Anxieties	0·4778	−0·4614	−0·2731
Obsessionals	0·2240	0·2714	−0·4364
Depressives	0·1034	−0·8979	0·3266
Hysterics	0·1237	−0·0802	0·6707
Psychopaths	0·4469	1·0603	0·0889
Mixed states	0·1317	−0·0598	−0·1671
Normals	−1·5677	0·0859	−0·0650

The co-ordinates of the group means in these three dimensions are given in the second part of the table, and their scatter is shown graphically in Figs. 8.2, 8.3 and 8.4.

On the whole the relationship between the groups shown in these graphs is consistent with the psychological theories of Eysenck and Hildebrand. In the first dimension the normal group is contrasted with the neurotic groups, all huddled together at the opposite end of the scale. This is seen equally clearly in Figs. 8.2 and 8.3. Just such a relationship is ascribed by them to neuroticism. In the second and third dimensions, as Fig. 8.4 shows, the anxiety states and the hysterias are well separated. A line passing through the two mean points would be nearly parallel to the axis of the third dimension, and would exhibit a relationship between the group means of the kind they attribute to I–E.

The surprising feature of the dispersion is thus its second axis, which arrays the psychopaths and the depressives at opposite poles and the normal cases midway between them, with most of the neurotics on the depressive side of the normals. For this result no prior hypothesis was advanced and no special selection of tests was made. It is perhaps all the more interesting on that account.

A possibility worth considering is that phenomena at present being ascribed to I–E are really heterogeneous and should be ascribed to two independent factors. In his theoretical formulation, quoted in section 2 of Chapter 8 above, Eysenck explicitly associates hysteric with psychopathic disorders on the one hand, locating them at the extraverted end of the I–E continuum, and implicitly associates reactive depressions with anxiety states on the other as dysthymic, locating them at the introverted end. Fig. 8.4 shows two independent axes, one running from

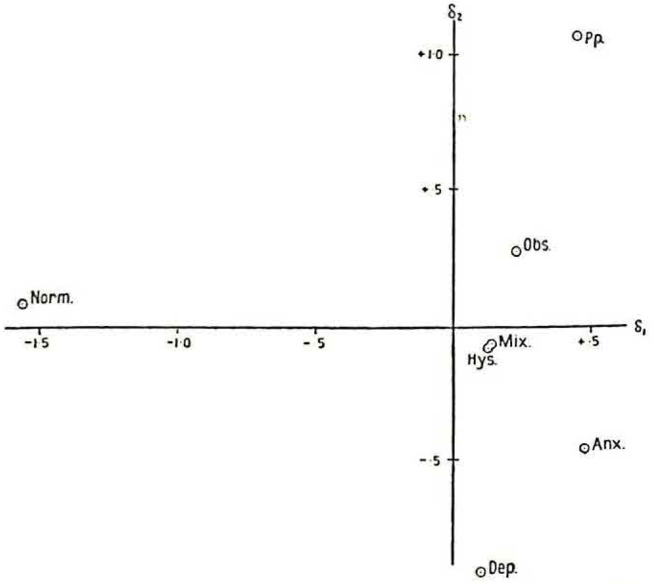


FIG. 8.2. The Dispersion between Groups on its First and Second Vectors in the 10-Test Battery.

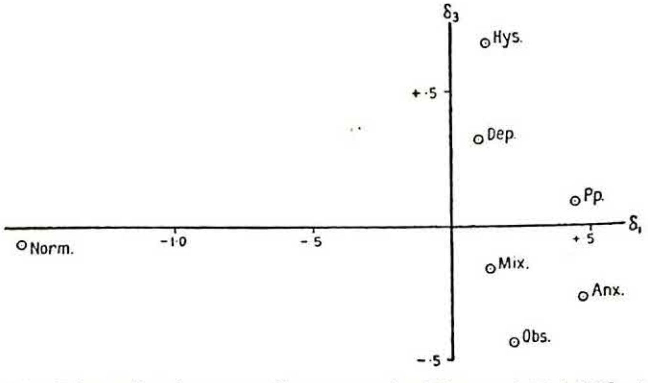


FIG. 8.3. The Dispersion between Groups on its First and Third Vectors in the 10-Test Battery.

anxiety across to hysteria and the other from psychopathy to depression. Thus it illustrates the relationships he anticipated but expands I–E into a two-dimensional continuum.

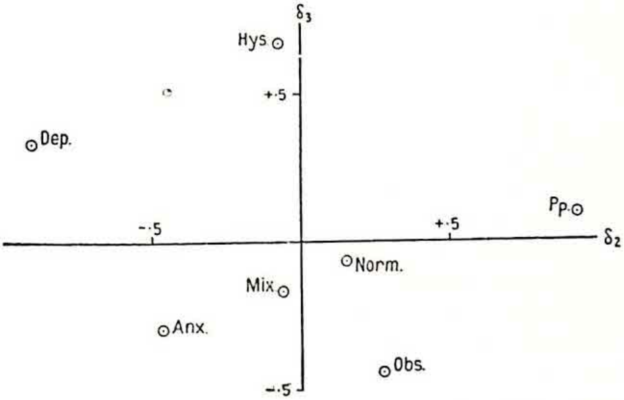


FIG. 8.4. The Dispersion between Groups on its Second and Third Vectors in the 10-Test Battery.

5. THE PROJECTION OF A D-SPACE ON AN f-SPACE

In the previous section the vectors derived from the discriminatory analysis were compared with those derived by Hildebrand from his factor analysis. The discriminating power of the factor analysis proved, not surprisingly, inferior to that of the canonical analysis. The latter technique is developed specifically for maximizing discriminance, while the former, as I have explained, is designed and was employed for a different purpose. We may however enquire why the deficit is as large as it is.

In a limited sense the canonical analysis is an optimal procedure, and extracts all the information obtainable from the data about the differences between the groups. Assuming that the multi-dimensional dispersion within groups is normal and homogeneous all the relevant information is contained in UT^{-1}. All the differences between the groups discoverable in the m-space are exhibited in the D-space. Extracting the principal components of the dispersion there completes a search of the D-space and therefore also of the m-space.

But precisely the same results could have been obtained if T had been resolved into its principal components first. Instead of being expressed directly as derived functions of the test scores the canonical axes would then have been expressed as derived functions of the principal components, themselves derived functions of the test scores. The procedure would have been unnecessarily circuitous and laborious but could have been carried through without any sacrifice of accuracy, and a complete search would have been made.

Accuracy is lost inevitably at two stages in the course of a factor analysis of the kind employed by Hildebrand. The first loss occurs when the common factors are extracted for consideration and the specific factor variance is discarded as negligible. The assumption that at some stage the residuals become insignificant is so widely prevalent that no one can be blamed for accepting it uncritically. Nevertheless it is open to criticism. For instance Kendall's comment (1957) is pertinent: 'No non-zero λ requires any test of significance in this sense: a sample with *d* effective dimensions cannot arise from a population with fewer than *d* dimensions. However small a value of λ may be, therefore, it cannot arise from a parent value of zero. It may be negligible but it is always significant.' I have published examples (1953) where the size of the smallest component of a dispersion is subject to experimental control and cannot be neglected. And the specific variance Hildebrand excludes from consideration is not just a small proportion; according to his figures it is 68% of the total.

If this were all, the search would have been restricted to the *f*-space defined by the unrotated vectors of the first four centroid factors, and the part of the D-space confined within this *f*-space could have been explored in its entirety. But a further loss is incurred when rotation is carried out and two of the rotated vectors are discarded as measuring traits lacking interest from the point of view of the psychological hypothesis. Search is consequently restricted to a reduced *f*-space of two dimensions, and only the cross-section of the D-space confined within it is discernible.

The results would be interesting if loss of this second kind could be excluded by some method of analysis, and loss of the first kind, which is incurred inevitably when search is restricted to any *f*-dimensional subspace in an *m*-dimensional continuum, could be isolated and measured. But I cannot suggest any comprehensive method of analysis which will include all the different *f*-spaces discoverable by different factor-analytic procedures for the same number of factors. We cannot for instance argue that all such *f*-spaces must be contained within the space defined by the first *f* principal components, although it includes a larger proportion of the total dispersion than any other *f*-space. For if the major axis of the dispersion passes through a point P at some distance from the origin, other points, say P′ for instance, can be found in the near neighbourhood of P such that the vector OP′ projects onto any or every other principal axis.

But it is not difficult to compare the variance between groups (ν_l, say) with the variance within groups, i.e. λ, along a principal axis of the standardized dispersion within groups. Thus taking each axis in turn we can discover whether the ratio $\nu_l : \lambda$ tends to be greater when λ is large, or whether some other relationship exists or no obvious relationship at all. This subjects to scrutiny the postulate that the dimensions where differences between groups are most significant are those where members of the same group vary most widely from one another. This postulate underlies the use of factor analysis here, where it is supposed that neuroticism and I–E describe major dimensions of variation between individuals and also account for the incidence of different forms of mental disorder. And analogous postulates underly many other applications of factor analysis as already explained in Chapter 6, section 3.

The latent roots and vectors of R_t, the matrix of correlations given in Table 8.2, are set out in full in Table 8.6.[1] Each latent root, λ, defines the variance along the vector l, which specifies one of the orthogonal axes of R_t. Thus their total, $\Sigma\lambda = m$, defines the total variance of the standardized dispersion within groups in all its dimensions. The ratio $\lambda/\Sigma\lambda$ gives the proportion contributed by each latent root to this total. It is listed as a percentage in the centre column below, and serves to show how the latent roots diminish progressively in size. Another total is formed by adding the variance between groups along the axis l, i.e. the quantity here denoted v_l, to the variance within groups, viz: $\lambda + v_l = V_l$. Thus the ratio v_l/V_l shows the proportion of the total variance along each axis of R_t due to the differences between group means in that dimension. I have preferred it to the ratio v_l/λ for descriptive purposes and list it, also as a percentage, in the right-hand column below.

	Ratio	
Component	$100\lambda/\Sigma\lambda$	$100\ v_l/V_l$
First	29·89	17·43**
Second	16·34	10·21**
Third	11·62	4·06
Fourth	9·47	10·64**
Fifth	8·64	5·68
Sixth	6·45	14·36**
Seventh	6·11	3·16
Eighth	5·20	11·61**
Ninth	3·82	18·97**
Tenth	2·46	8·30*

* Significant at the 5% level
** ,, ,, ,, 1% ,,

---

[1] From every point of view R_t seems preferable to R_s for use here (cf. section 1, Chapter 7, for definitions). Mathematically because it is directly related to T and hence to UT^{-1}. Descriptively because it gives the best estimate of the dispersion in the normal population. A sample composed of 15% hysterics, 6% anxiety states . . . 13% normals cannot be regarded as representative of any population whatever. But there are no obvious objections to assuming initially that the 22 E-controls are a representative sample of normal adult men; and since this group is homogeneous with the others in its internal variation, as shown by the s.d.'s given for each group separately on each test in Hildebrand's thesis, summation within groups is reasonable. Hildebrand was influenced by similar considerations; he attempted to approximate to R_t by excluding the supposedly extreme groups from his factor analysis.

TABLE 8.6

Latent roots and vectors of the matrix of correlations in Table 8.2

			Principal components:		
	First	*Second*	*Third*	*Fourth*	*Fifth*
Roots	2·9894	1·6340	1·1617	0·9468	0·8639
Variables			Vectors (normalized)		
1.	−0·0377	−0·3264	0·6581	−0·0949	0·4255
2.	0·4415	0·0015	−0·2716	0·0318	−0·1283
3.	−0·2414	−0·0944	−0·3227	0·5158	0·6393
4.	0·3474	0·0227	−0·2659	−0·3831	0·3312
5.	0·4198	−0·2150	0·0303	0·3729	−0·0107
6.	0·3933	−0·0595	0·3215	−0·1644	0·2662
7.	0·4839	0·0852	−0·1131	−0·0137	0·1379
8.	−0·1913	0·4682	−0·1113	−0·4791	0·3308
9.	0·1129	0·5699	0·0915	0·3716	0·2319
10.	0·1016	0·5319	0·4249	0·2078	−0·1755
	Sixth	*Seventh*	*Eighth*	*Ninth*	*Tenth*
Roots	0·6450	0·6109	0·5204	0·3816	0·2463
Variables			Vectors (normalized)		
1.	0·3128	−0·3025	0·2060	−0·1362	0·1270
2.	0·3941	−0·0971	0·3818	−0·5106	−0·3778
3.	−0·0346	0·2685	0·2692	0·0136	−0·0780
4.	−0·4784	−0·4061	0·3115	0·2276	0·0674
5.	0·2465	−0·1926	−0·2376	0·6334	−0·2848
6.	−0·3402	0·5083	−0·2698	−0·1812	−0·3923
7.	0·2902	0·3376	−0·0736	0·0131	0·7209
8.	0·4858	0·0836	−0·1224	0·2488	−0·2696
9.	−0·1093	−0·4278	−0·3981	−0·3187	0·0491
10.	−0·0905	0·1893	0·5796	0·2666	−0·0274

Significant differences between the group means seem on this evidence equally likely to occur in dimensions where individual variation about the means is large or small. The greatest occurs in the ninth component and the next greatest in the first. Others are scattered in between instead of clustering among the larger components as, according to the postulate, they should.

The finding is all the more remarkable because the postulate impugned seems otherwise so acceptable in the present context. The groups are segregated, in the opinion of the psychiatrists, by natural causes operating on constitutionally different individuals. The tests are of two kinds, chosen to replicate one another in measuring the two sorts of constitutional difference supposed to be most important diagnostically. These considerations would seem adequate to encourage the belief that the best sub-space to search for evidence of differences between the groups is the one where individual variation is widest, i.e. the sub-space defined by the first two, or perhaps three components. But, as it now appears, a search started at the opposite end, where individual variation is smallest, might have proved equally rewarding.

Knowing, as we do now, that the D-space is three dimensional, we may go on to enquire how much of the evidence of the variation in its three dimensions, δ_1, δ_2 and δ_3, is provided by each of the principal components of R_t. The correlations between the two sets of components show the relationship conveniently:

Correlations of the principal components of the D-space with those of R_t	1	2	3
λ_1	0·544	0·193	−0·275
λ_2	−0·007	0·374	0·562
λ_3	−0·153	0·280	−0·159
λ_4	0·016	−0·680	−0·050
λ_5	0·075	0·366	0·081
λ_6	0·514	−0·049	−0·134
λ_7	0·097	−0·291	0·029
λ_8	0·207	−0·234	0·687
λ_9	−0·597	−0·047	−0·064
λ_{10}	−0·031	0·058	−0·278

Evidence of variation in δ_1 is to be found in λ_1, λ_6 and λ_9 to an almost equal extent; δ_2 mainly in λ_4, though λ_2 and λ_5 contribute to a lesser extent; δ_3 mainly in λ_8, but λ_2 is also relevant. Thus to obtain a good projection of the entire D-space all the principal axes of R_t marked ** on p. 289 would need to be explored; and perhaps others.

But even if it is not safe to assume that the major dimensions of R_t are those where search for discriminance is most likely to be rewarded, something short of a complete analysis may still be sufficient to discover something about the canonical variates, and in particular, *how many* discriminate significantly between the groups. Assuming only that some evidence may be found in any of the principal axes of R_t, a sufficient amount may be accumulated from a limited number. And it will still be reasonable to consider the largest components first, as they are the first to be extracted. How many of the principal axes of R_t, beginning with the largest, have to be investigated to obtain sufficient evidence to prove that this D-space is 3-dimensional?

A search restricted to the 3-space of λ_1, λ_2 and λ_3 turned out to be insufficient. But a peculiarly instructive picture is obtained. The three canonical variates have the latent roots, for which I will retain the same notation,

$\delta_1 = 0{\cdot}2163$, hence $\chi^2 = 34{\cdot}61$ with d.f. 8**
$\delta_2 = 0{\cdot}1133$, ,, ,, $= 18{\cdot}12$,, ,, 6*
$\delta_3 = 0{\cdot}0375$,, ,, $= 6{\cdot}00$,, ,, 4

The normalized vectors of the first two are

	δ_1	δ_2
λ_1	0·9840	0·0706
λ_2	−0·0685	0·9975
λ_3	−0·1645	0·0074

Thus the latent vector of δ_1 coincides almost exactly with that of λ_1, and δ_2 with λ_2.

That is to say, the D-space appears only in a 2-dimensional cross-section; the first canonical variate coincides almost exactly with the major axis of the dispersion of individuals about their group means; and the second variate almost exactly with the next principal axis of that dispersion. The expectations deduced from the hypothesis this experiment was designed to test are supported admirably in every respect by the evidence shown in this picture, cf. Fig. 8.5, though they are not sustained by the evidence as a whole. The projections of the group means on λ_1 and λ_2 are given below:

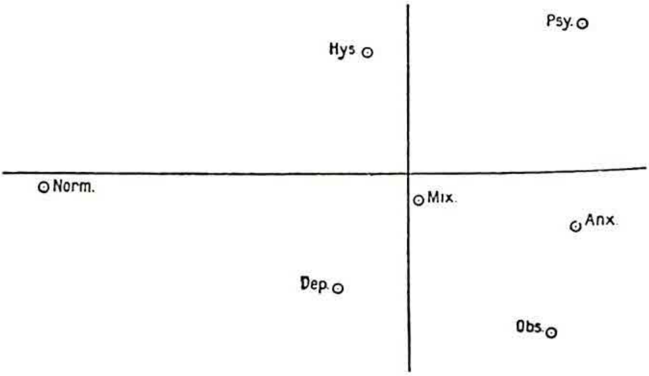


FIG. 8.5. The Scatter of the Groups in the Plane defined by the first two Principal Components of the Dispersion within Groups.

Group	λ_1	λ_2
Anxieties	5·9102	0·9100
Obsessions	5·8247	0·3839
Depressions	4·7539	0·6976
Hysterias	4·9635	1·8174
Psychopathics	6·0831	1·8432
Mixed neuroses	5·1883	1·0929
E-controls	3·4712	1·2602
All cases	5·1557	1·2198

The third dimension of the dispersion between groups first appears above our horizon when we start exploring the 4-space of λ_1 to λ_4. The latent roots of the canonical variates are augmented to

$$\begin{array}{lllll}
\delta_1 = 0{\cdot}2189, & \text{hence } \chi^2 = 35{\cdot}028 & \text{with } 9 & \text{d.f.} & ** \\
\delta_2 = 0{\cdot}1886, & \text{,,} \quad \text{,,} \; = 30{\cdot}169 & \text{,,} \quad 7 & \text{,,} & ** \\
\delta_3 = 0{\cdot}0750, & \text{,,} \quad \text{,,} \; = 11{\cdot}993 & \text{,,} \quad 5 & \text{,,} & * \\
\delta_4 = 0{\cdot}0038, & \text{,,} \quad \text{,,} \; = 0{\cdot}603 & \text{,,} \quad 3 & \text{,,} &
\end{array}$$

The normalized vectors for the first three are

	δ_1	δ_2	δ_3
λ_1	0·9449	0·1892	−0·0102
λ_2	0·0736	−0·6240	0·7189
λ_3	−0·0916	−0·2820	−0·5556
λ_4	−0·2254	0·7038	0·4176

The vector for δ_1 does not yet appear to diverge far from that of λ_1, but those for δ_2 and δ_3 no longer tend to converge with any of the principal axes of R_t.

One conclusion can be drawn confidently from this investigation of the projection of a D-space on an f-space, where $f < m$; it is this: if the projection is p-dimensional the D-space must be p-dimensional *at least*. Canonical analysis cannot be used restrictively as Spearman and the early factorists used factor analysis: p is a minimum, not a maximum.

6. STORMS' TREATMENT OF THE DATA

Very little theoretical justification has been found here for expecting the vectors defined by a factor analysis to coincide with those defined by a discriminatory analysis of the same data; and the evidence examined has shown that they do not converge closely even under particularly favourable conditions. The reason is not because one method of analysis is better or worse than the other, but because the two methods of analysis are designed to serve different purposes.

And for the same reason it is arguable that the guiding principles to be followed in assembling a set of tests for a factor analysis are not the best ones for an experiment in discrimination. For a factor analysis starting with prior expectations that one psychological trait, F1, is observable in a certain range of test situations, another trait F2, in another range, etc., enough tests are required to cover the range for F1, and enough more to cover the range for F2; and in the set finally assembled there will be a large number of tests clustering together in a few distinct categories.

For a discriminatory analysis, on the other hand, the tests chosen for inclusion in the battery should serve either as *predictors* or *suppressors*. A predictor is any measurement which is found to correlate positively or negatively with the criterion. And the predictors assembled in a battery should be as far as possible uncorrelated with one another. For if they are their evidence is cumulative, but if on the contrary they are correlated what can be inferred from one may add nothing to what is already known from another. A suppressor is one which correlates positively or negatively with a predictor but not with the criterion. Its use is indirect and perhaps best explained with the aid of an example: To measure verbal ability independently of intelligence is difficult with one test alone, for tests of verbal ability usually reflect intelligence to some extent. But tests of intelligence which do not involve verbal ability are available; and a weighted summation of two tests, one of verbal and one of non-verbal intelligence may provide the required measurement. The non-verbal test is used as a suppressor, and scores on it receive a negative weight. In general we may say that suppressors control components of the total variance of the predictors which would otherwise contribute to the error variance of the estimate of the criterion. And

TABLE 8.7

Data used by Storms

PART 1. Means of groups (see footnote) on tests:

Group:	Suggestibility Reversals	Leg Persistence
Anxiety states	3·3656	1·6234
Obsessions	3·7611	1·6443
Depressions	3·0700	1·5706
Hysterics	4·6244	1·3046
Psychopathies	3·0368	1·6818
Mixed neuroses	3·3726	1·6029

PART 2. S.d.'s within groups of tests enumerated below:

1.	120·8664
2.	7·7161
3.	2·4372
4.	0·1113
5.	47·7987
6.	36·7880

PART 3. Correlations within groups, R_t. Tests 4, 5 and 6 have been reversed, i.e. scored negatively:

	2	3	4	5	6
1. R Scale	0·02626	0·03460	0·00143	0·26072	0·09598
2. Lie scale		−0·17471	0·15891	0·29381	0·22378
3. Suggestibility reversals			0·17490	0·01383	0·12029
4. Leg persistence				0·10033	−0·06931
5. MMQ					0·46466
6. Annoyances					

PART 4. L^{-1}, when $LL' = R_t$:

	1	2	3	4	5	6
1.	1·00000					
2.	−0·02627	1·00034				
3.	−0·03986	0·17863	1·01643			
4.	0·01135	−0·20279	−0·21696	1·03569		
5.	−0·27361	−0·31362	−0·05165	−0·04992	1·08817	
6.	0·03387	−0·17665	−0·19981	0·19591	−0·51718	1·16937

Note: Group means on other tests are to be found in Table 8.3.

again, there are no advantages in including a large number of suppressors which replicate one another. In the set finally assembled for the discriminatory analysis there will be a fairly small number of tests which do not tend to form clusters, except where suppressors are linked with predictors.

Storms was guided by considerations of this kind in selecting a battery from Hildebrand's tests for his exercise in discriminatory analysis. He limited the measures of neuroticism and I–E to four: the Maudsley Medical questionnaire and lie scale, Annoyances and the R scale from the Guilford-Martin inventory. To these he added two tests which revealed significant differences between neurotic groups but had negligible factor loadings on neuroticism and I–E in Hildebrand's analysis (cf. Table 8.1). They were the Suggestibility Reversals and Leg Persistence tests. A square-root transformation was used for the former and a logarithmic transformation for the latter, as the crude scores on both tests have very skew distributions.

The data relate to 145 cases only, as these two tests were not given to the E-controls and the H-controls were regarded as suspect. The essential particulars are given in Table 8.7, and the triangular matrix L^{-1} is included for the convenience of anyone interested in reconstructing the complete analysis.

Although only six tests are included, the battery is distinctly more sensitive to diagnostic differences than the ten-test battery discussed in the preceding sections (2 to 5), which was originally assembled with factor analysis in mind. Some details of the discriminant functions are given by Storms. The summary below can be compared with the summary for the ten-test battery in section 3.

	Obs.	Dep.	Hys.	Psy.	Mix.
Anxiety states	—	—	++	++	—
Obsessions		—	+	—	—
Depressions			+	++	—
Hysterias				++	++
Psychopathies					+
Mixed neuroses					

Note: +, discrimination significant at the 5% level
++, ,, ,, ,, ,, 1% ,,

The means for the suggestibility reversals and leg persistence tests given in part I of Table 8.7 reveal conspicuous differences between the hysterics and the other patients, and the discriminant functions isolate them effectively. The isolation of the psychopathic group is also increased; in consequence the difference between the hysterics and the psychopaths is the widest found in any of the comparisons. But the anxiety states, mixed neuroses, obsessions and depressions remain confused together, so there is still scope for further refinements.

The first three latent roots from the canonical analysis are

$$\begin{aligned}\delta_1 &= 0{\cdot}2674, \text{ hence } \chi^2 = 37{\cdot}169 \text{ with } 10 \text{ d.f.}^{**}\\ \delta_2 &= 0{\cdot}2357, \quad ,, \quad ,, = 32{\cdot}760 \quad ,, \quad 8 \quad ,,\ ^{**}\\ \delta_3 &= 0{\cdot}0595, \quad ,, \quad ,, = 8{\cdot}269 \quad ,, \quad 6 \quad ,,\end{aligned}$$

The third is insignificant. There is in fact no variate to correspond with the first obtained from the ten-test battery, which contrasted the normals with the neurotics but showed no differences between the neurotic groups. The disappearance of this vector is not surprising, seeing that no control group has been included. The first and second vectors from the six-test battery are to be compared with the second and third from the ten-test battery.

The vectors of the two significant canonical variates are

Test	δ_1	δ_2
1. R scale	0·3349	−0·6756
2. Lie scale	0·3953	0·4033
3. Suggestibility reversals	0·6355	0·0779
4. Leg persistence	0·5429	0·1274
5. MMQ	−0·0539	−0·1800
6. Annoyances	0·1733	0·5711

Note: The signs given are appropriate for use with reversed scores on tests 4, 5 and 6.

The distribution of the group means in the two dimensions is shown in Fig. 8.6.

There is a striking resemblance between the δ_2 vector above and the vector of the second canonical variate from the ten-test battery. The distributions of the group means along the two axes are almost identical. This can be seen by a comparison of Figs. 8.4 and 8.6: in both dimensions the psychopathies lie at the opposite end to the depressions and the anxiety states, while the remaining groups cluster around the general mean.

This leaves δ_1 above presumably comparable to some extent with δ_3 in the ten-test battery: since the 6-space and the 10-space have four dimensions in common, where the points locating individuals are identical, the alternative hypothesis that all resemblances are purely coincidental seems rather far-fetched. The introduction of the suggestibility reversals and leg persistence tests has extended the variance in this dimension (it has the smallest of the three significant latent roots in the ten-test battery and the larger of the two in the six-test) by widening the distance between the hysterics and the remaining groups. As a result it is less easy to find any vector with properties such as are ascribed to I–E in the plane exhibited in Fig. 8.6 than it is in Fig. 8.4. The axis of δ_1 in Fig. 8.6 simply provides a dichotomy separating the hysterics from the other groups.

This is not a summary of Storms' paper. Much of the evidence he quotes is taken from the analysis of the ten-test battery reported in previous sections. And the evidence I have given here relating to his six-test battery does not appear in his paper. Naturally I hope that my presentation makes clearer where and how the differences between the results of the factor analysis and the canonical variate analysis have arisen.

Storms uses this evidence for what might almost be described as an attack on Eysenck, which is not quite justified in my opinion. Eysenck can hardly be considered personally to blame because factor analysis and discriminatory analysis serve different purposes, benefit most from differently designed experiments, and lead to different results. He is none the less free to postulate that introversion-extraversion is a psychological trait relevant to problems of psychiatric diagnosis, to look for a method of measuring it by the weighted summation of scores on specified objective tests, and to employ factor analysis to find suitable weights. I am sure that many eminent psychologists would concur that this is a very reasonable line of research.

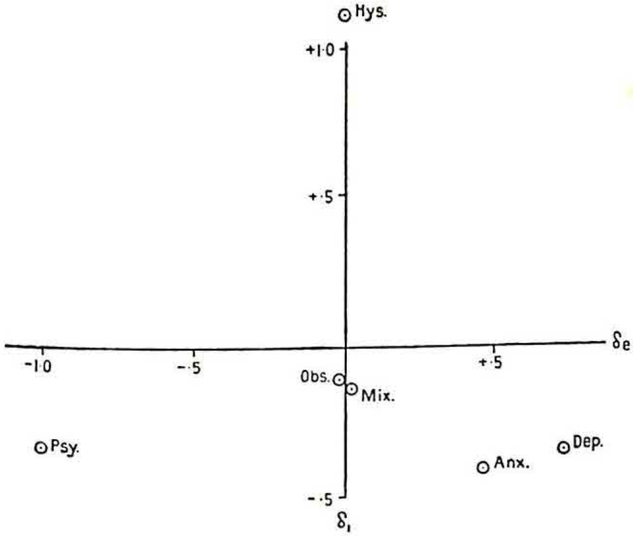


FIG. 8.6. The 2-dimensional Dispersion between Groups found in the 6-Test Battery.

But Storms is fully supported by the evidence when he emphasizes that risks may be incurred if hysterics are pooled with psychopaths to form a group of neurotic extraverts for experimental purposes. Such a group might be homogeneous with regard to I–E but heterogeneous in other respects; and without knowing a good deal about the experiment under consideration it would be difficult to say whether the risks are likely to be serious or negligible in the particular context. Similar risks may also arise if cases of anxiety states and reactive depressions are pooled to form a group of dysthymics or neurotic introverts. Although Hildebrand's evidence is not sufficient to demonstrate any significant difference between these diagnostic categories the distance between their group means and their locations on the axes in Fig. 8.4 are enough to suggest that further research may lead to effective discrimination with objective tests. Other evidence of differences in test results (cf. Foulds, 1951 and 1953) can be cited in corroboration.

7. SUMMARY

Canonical analysis is applicable to the measurements obtained by giving a number of psychological tests to individuals belonging to several groups. The means of the *g* groups form a scatter of *g* points in the *m*-space defined by the *m* tests, which is the dispersion between groups. The individuals' scores form scatters about their group means which can be collected to form a dispersion within groups. When the two dispersions are contracted simultaneously to the extent necessary to normalize the variance within groups in all dimensions, the between-group dispersion assumes an ellipsoidal form in a D-space which may be *m*-dimensional or less. Its major axis is the dimension where the most significant differences between the groups are to be found, the next the dimension where differences independent of the first are most significant, etc. The canonical analysis of discriminance is a principal component analysis applied to this contracted dispersion between groups to define its axes and test the significance of the variation there. Working methods are described and two examples are discussed.

The axes or dimensions so defined do not necessarily coincide with those found by factor analysis of the same data. There are no *a priori* reasons for expecting the most significant differences to appear in the dimensions where the variance of the individuals about their group means (or about the general mean for all groups) is widest. In an experiment where the conditions were particularly favourable for the appearance of a relationship, no evidence of any relationship was found between the amount of variation within groups in a given dimension and the significance of the difference between groups there.

If factor scores are calculated for *f* factors an *f*-dimensional sub-space is defined within the *m*-space, and a canonical analysis can be made of the projection of the dispersion between groups to be found there. This will give some incomplete information about the principal axes of the original dispersion in the D-space. As *f* increases towards *m* the information becomes more complete. The evidence examined here does not suggest that there are any advantages in considering the *f*-space where the dispersion within groups is widest except that it is the most convenient one to define first. But there may be special reasons for supposing that there are particular advantages in some psychological experiments, which have not been considered here.

The principles to be followed in selecting a battery of tests for a canonical analysis are not the same as for a factor analysis.

8. BIBLIOGRAPHY, REFERENCES AND ACKNOWLEDGEMENTS

Canonical variate analysis, the term which has come into use among psychologists for the kind of analysis described here, has too wide a generality to be quite appropriate. Canonical analysis of discriminance is my own suggestion: on the one hand the problem of discrimination is what determines the form the canonical analysis takes, and on the other canonical analysis is only one of the ways of treating the problem of discrimination. Together, the two terms 'canonical' and 'discriminance' seem to specify this kind of analysis precisely.

Rao, whose presentation of the subject is authoritative (1952), discusses it under the headings of classificatory problems and the generalized distance between groups. Kendall deals with it in his chapter on discriminatory analysis (1957) as the case of k populations. Both authors make use of the example published by Rao and Slater (1949) which seems to have become the standard one. Kendall's account of the historical development of methods for analysing multidimensional variations between groups, and his systematic treatment of problems of testing significance will, I believe, be found most helpful. He shows the historical importance of the original contributions of Mahalanobis (1936, 1948) and Wilks (1932). Matrix algebra is usually introduced into mathematical courses at the advanced level, but as it is a development of methods of solving simultaneous equations the elementary operations it employs can be explained to anyone who has encountered such equations in algebra. Aitken explains them in this way in his excellent little textbook (1939). For wider information in this field Lanczos (1957) may be recommended. Mr. J. G. Hayes suggested using the triangular matrix L to simplify computations (cf. section 3, Chapter 7). Further details of this working method are given by Fox (1950).

My thanks are due to Professor H. J. Eysenck for the facilities and encouragement he has given me, and to Miss N. Hemsley who was of very great assistance with the computations.

Epilogue

9

THE PLACE OF THEORY IN PSYCHOLOGY

H. J. Eysenck

PSYCHOLOGISTS hold quite widely differing views about the usefulness and the general role of theory in psychology. Some stress simple inductive methods exclusively, maintaining that the time is not yet—or perhaps may never be!—when more ambitious generalizations of a theoretical kind would be appropriate. Others emphasize the similarities between all sciences, and conclude that 'what is good for physics must be good for psychology!' In all the discussion that has taken place, sight is often lost of a very simple fact which, to my mind, is quite crucial. Scientific theories differ among themselves to such an extent that any discussion about the place of 'theory' in science should really be restyled: a discussion about the place of *theories* in science. In particular, there is a continuum ranging all the way from *weak* to *strong* theories, and failure to pay attention to the distinction between these two kinds of theories renders much discussion, and many criticisms, quite meaningless.

The theory which is always quoted as the perfect example of the application of the hypothetico-deductive method is of course Newton's theory of gravitation. This is a good example of a *strong* theory in science, and it may be worth while to look at those features of it which make it so. In the first place, it is based on a very large number of accurate observations, made over many years by large groups of people. In the second place, it brings together a number of sub-fields in which quantitative laws—such as those of Kepler and Galileo—had already been discovered and verified. In the third place, the phenomena in question were relatively clear-cut and unambiguous; in particular, they were not embedded in or entangled with groups of other phenomena. In the fourth place, the mathematical relations in question were not of a very complex order—compared, say, with modern atomic theory. In the fifth place, and largely as a consequence of the preceding points, predictions were uncommonly straightforward and precise; verification and infirmation of deductions did not give rise to special problems.

While for all these reasons Newton's theory is a good example for the beginner, it is seriously misleading for the practical scientist. Very few scientific theories are in fact *strong* theories of this kind; most of them lie toward the opposite end of a continuum going from strong to weak. The typical weak theory in science shows all the opposite characteristics to those mentioned in connection with Newton's. Only few observations, and these of doubtful accuracy, are available. Few quantitative or even qualitative laws, universally established, are available in sub-fields. The nature of the phenomena in question is by no means clear-cut or well understood. Mathematical relations are often very complex, and predictions are neither straightforward nor precise. What, it may be asked, is the use of such a weak theory?

The answer may be given in the words of the famous physicist J. J. Thomson: 'A theory in science is a policy rather than a creed.' I quoted these words on the front page of my book *Dynamics of Anxiety and Hysteria* to indicate the heuristic nature of the theory there developed, and I believe that they contain the key to a proper understanding of the function of theory in psychology—where nearly all theories are weak theories almost by definition. The value of a weak theory, to put it briefly, lies in the fact that *it directs attention to those problems which most repay study from a systematic point of view*; in Thomson's words, it defines a policy of action and research. It is by giving rise to worthwhile research, rather than by necessarily being right, that a weak theory makes its greatest contribution to science.

An example or two may make clearer what I have in mind. On the basis of the theory developed in *Dynamics of Anxiety and Hysteria*, I predicted that extraverted people would have greater reminiscence effects after massed practice than would introverted people. Using the pursuit rotor, several investigators have tested this deduction. All those using a practice period of five minutes have verified the deduction; none of those using a practice period of ninety seconds have done so. This illustrates two points. In the first place, our knowledge of the sub-field of reminiscence was too restricted to allow of a more precise specification of differences between extraverts and introverts; clearly length of practice should have been specified as an important parameter, but could not be so specified because of lack of knowledge in this respect. In the second place, the partial verification of the deduction is of considerable interest; it suggests relationships between personality and learning theory which are worthy of closer study. At the same time, the direction which such study should take are emphasized by the findings; we clearly must concentrate on time relations in the practice period, and presumably also in the rest period. In other words, the theory leads us to a more precise study of the growth and decay of reactive inhibition, which is supposed to underlie reminiscence; the fact that positive relations with extraversion have been firmly established *for certain time intervals* suggests that a policy of research concentrating on this aspect will not be a waste of time.

As another example, let us take two deductions from the general theory:

(1) Extraverts will show greater visual satiation;
(2) Extraverts will have greater difficulties in maintaining visual fixation.

At first sight, these two deductions may appear contradictory. Satiation can only be measured by the determination of figural after-effects following visual fixation of an inspection figure; failure to maintain fixation leads to lower figural after-effects. According to (1), therefore, extraverts should show greater figural after-effects, while according to (2) they should show lesser effects. Any outcome, including equality, could therefore apparently be 'explained' in terms of these two deductions. This, however, is not quite true. Fixation should not begin to deteriorate until some considerable time had elapsed, say 30 seconds; satiation should be almost immediate. A combined prediction would therefore be that for short inspection times extraverts would have greater figural after-effects, while for long inspection times their after-effect scores would be lower. According to this hypothesis, there should also be one point at which extraverts and introverts should be equal, namely at some intermediate length of inspection time. Again no precise prediction can be made as to the times involved because our knowledge of figural after-effects is not at the moment extensive or quantitative enough for such predictions. But again a weak theory has directed attention to an important area of research, and the positive results reported to date with different types of figural after-effect tests suggests that a policy of research concentrating on this complex of relationships would not necessarily fail to repay the investigator for his effort.

These considerations suggest that it is relatively absurd, particularly at an early stage of development of a scientific theory, to construct a score board, with each success and each failure of prediction written in, to give a kind of batting average. There are usually many reasons for the (apparent) failure of a prediction, but far fewer for its success; consequently failures are of much less interest than successes in evaluating a theory at an early stage. In particular, failures may arise, not because the theory is in error, but because the deduction made in a particular sub-field makes use of a theoretical model in that sub-field which is incorrect; this does not in any way invalidate the general theory. As an example, we may take the prediction that extraverts would show a more pronounced bowing in the serial position curve effect in nonsense syllable learning. According to the Hull-Lepley theory, this effect is due to inhibition of delay, and the hypothesis that extraverts are more prone than introverts to generate inhibitory potential mediates a clear-cut prediction. This prediction could not be verified, however, and neither could another one linking depressant drugs with an increased bowing effect. A special experiment was therefore carried out to test the Hull-Lepley theory, by comparing bowing effects with and without intervals between successive presentations of the nonsense syllable series; the differences which should have appeared according to the Hull-Lepley theory failed to materialize, and consequently it was concluded that the Hull-Lepley theory was itself in error (Eysenck, 1959c). Failure to verify the deduction from the general personality theory was due, therefore, not to an error in the theory itself, but to an error in that part of learning theory used to mediate a particular prediction.

Even when there is no such error, predictions may not be verifiable for a variety of reasons although the theory is in fact correct. Two well-known examples are the failure to observe parallax in stellar positions, which was one of the most direct predictions made from Copernicus's heliocentric theory of the planetary system, and the failure to discover the capillaries which according to Harvey's theory of the circulation of the blood should intervene between arteries and veins. The (correct) explanations given by Copernicus and Harvey, namely that the stars were too far away, and the capillaries too small, to make observation of the predicted effect possible with available instruments, were not at the time susceptible to proof. Scientists tended to accept the theories in question because they unified a large number of facts, although these apparently crucial deductions remained unverified for a long time.

Altogether, the notion of an *experimentum crucis* to decide the correctness of a theory, or to decide between alternative theories, is one which appears more frequently in the pages of popular expositions of scientific method than in actual practice. It will be remembered that two members of the Thomson family, father and son, were both awarded the Nobel prize in physics, one for showing conclusively that light was of the nature of a *particle*, the other that it was of the nature of a *wave*! Having performed crucial experiments to prove both of these alternative theories regarding the nature of light, physicists are still left with something which sometimes behaves like a wave, sometimes like a particle; they have also learned rather painfully that crucial experiments are seldom as crucial as they are supposed to be—even in connection with a strong theory!

We may put the whole matter slightly differently, by following a discussion given by Cohen and Nagel (1936). They take as their example Foucault's famous experiment in which he showed that light travels faster in air than in water. This was considered a crucial experiment to decide between two hypotheses: H_1, the hypothesis that light consists of very small particles travelling with enormous speeds, and H_2, the hypothesis that light is a form of wave motion. H_1 implies the proposition P_1 that the velocity of light in water is *greater* than in air, while H_2 implies the proposition P_2 that the velocity of light in water is *less* than in air. 'According to the doctrine of crucial experiments, the corpuscular hypothesis of light should have been banished to limbo once and for all.' However, as is well known, contemporary physics has revived the corpuscular theory in order to explain certain optical effects which cannot be explained by the wave theory. What went wrong?

As Cohen and Nagel point out, 'in order to deduce the proposition P_1 from H_1, and in order that we may be able to perform the experiment of Foucault, many *other* assumptions, K, must be made about the nature of light and the instruments we employ in measuring its velocity. Consequently, it is not the hypothesis H_1 alone which is being put to the test by the experiment—it is H_1 and K.' The logic of the crucial experiment may therefore be put in this fashion. If H_1 and K, then P_1; if now experiment shows P_1 to be false, then either H_1 is false *or* K (in part or complete) is false. 'If we have good grounds for believing that K is not false, H_1 is refuted by the experiment. Nevertheless the experiment really tests both H_1 *and* K. If in the interest of the coherence of our knowledge it is found necessary to revise the assumptions contained in K, the crucial experiment must be reinterpreted, and it need not then decide against H_1.'

We may now indicate the relevance of this discussion to our distinction between weak and strong theories. Strong theories are elaborated on the basis of a large, well founded and experimentally based set of assumptions, K, so that the results of new experiments are interpreted almost exclusively in terms of the light they throw on H_1, H_2 . . . H_n. Weak theories lack such a basis, and results of new experiments may be interpreted with almost equal ease as disproving H as disproving K (see the example given above of serial position learning effects). The relative importance of K can of course vary continuously, giving rise to a continuum; the use of the terms 'strong' and 'weak' is merely intended to refer to the extremes of this continuum, not to suggest the existence of two quite separate types of theories. In psychology, K is infinitely less strong than it is in physics, and consequently theories in psychology inevitably lie towards the weaker pole.

Weak theories in science, then, generate research the main function of which is to investigate certain problems which, but for the theory in question, would not have arisen in that particular form; their main purpose is not to generate predictions the main use of which is the direct verification or infirmation of the theory. This is not to say that such theories are not weakened if the majority of predictions made are infirmed; obviously there comes a point when investigators turn to more promising theories after consistent failure with a given hypothesis, however interesting it may be. My intention is merely to draw attention to the fact—which will surely be obvious to most scientifically-trained people—that both proof and failure of deductions from a scientific hypothesis are more complex than may appear at first sight, and that the simple-minded application of precepts derived from strong theories to a field like psychology (or modern sub-atomic physics) may be extremely misleading. Ultimately, as Conant has emphasized, scientific theories of any kind are not discarded because of failures of prediction, but only because a better theory has been advanced; this may account for the longevity of Hull's system in spite of the many onslaughts on it in recent years.

There is a further characteristic of weak theories, as contrasted with strong, which deserves mention. In strong theories the different postulates are *interdependent*; it is not possible to change one without changing the rest, and indeed throwing overboard the whole theory. In weak theories, such interdependence is much less marked, and changes in one part of the theory are quite permissible without the necessity of altering other parts as well. Thus Hull's theory of inhibition is peripheral and 'work'-oriented; I have preferred (for various experimental reasons) to work with a central theory of inhibition rather more akin to Pavlov's. This substitution, as well as many others, can be made without extending the framework of Hull's theory unduly; nothing of this kind would have been possible with Newton's theory of gravitation. Weak theories are very flexible; that is why they are such good guides for research; strong theories have an air of 'take it or leave it' which makes them superior as guides to action, but also less likely to lead to important new discoveries.

There is, of course, one point which all scientific theories have in common, and which decisively sets them off from non-scientific theories. However weak a theory may be, it must generate predictions which admit of experimental investigation. In other words, the theory must be *reality-oriented*, and the manipulations of reality implicit in its testing must be capable of being made explicit without ambiguity. The hypothesis that planetary motions can be explained in terms of angels pushing the planets around on their courses, preordained by God, is not a scientific theory because it suggests no direct experimental investigation. It is an interesting question whether such concepts as the Oedipus complex, or the super ego, or the archetype, have characteristics which make them suitable for use in scientific theories, however weak, or whether they are outside science altogether; it would take us too far to discuss this point in detail.

Some scientific concepts, of course, have properties which are very difficult to disentangle from the matrix of events, even in the laboratory. Take for example the notion of the *neutrino*, first suggested by Pauli in 1931 in order to explain certain facts of radioactivity; this particle not only has no electric charge, but also has only negligible mass when at rest. All the properties assigned to it were those needed to balance certain physical equations, and but for the development of atomic energy its actual experimental demonstration would have remained impossible. As W. F. G. Swann said, 'it has no act', i.e. there is nothing (or at least very little) in its performance which could be directly tested. Nevertheless, the neutrino has a very slight probability of interacting with any proton which it might encounter (one chance in 10^{18} roughly), and this has been used to demonstrate its existence. Hull's 'involuntary rest pauses' may give us just as much trouble as did the neutrino to physicists, and for the same reasons—they don't seem to have much of an act!

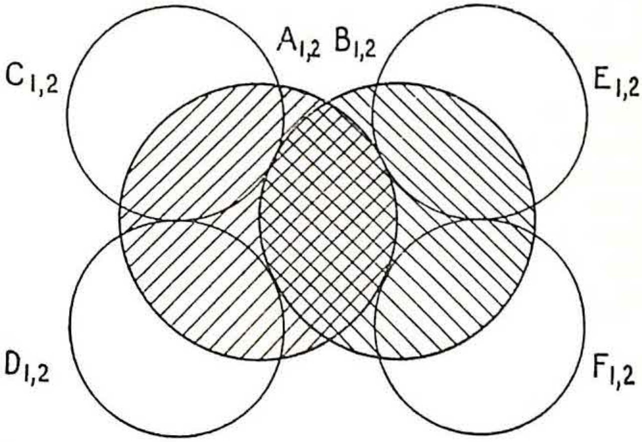
One of the main consequences of having to deal with weak rather than with strong theories is that attention should shift from considerations of right and wrong to considerations of fruitful or useless. It is difficult enough to disprove a strong theory; it is almost impossible to disprove a weak one. The point has been well made by Roby (1959): 'Ultimately, theory must answer to the facts, but this is not the only requirement placed on a theoretical system. Logical consistency, economy of assumptions, and even a degree of elegance are by no means secondary factors in determining the overall staying power of a theory. These patrician qualities are quite unlikely to mature, however, if the demand for direct descriptive capability is too insistent. It is not . . . contended that theory construction should be totally unresponsible to the general body of knowledge about behaviour. Rather it is held that point-by-point testing of isolated facets of a theory against specific behavioural phenomena or experimental findings is at odds with the whole purpose of theoretical abstraction. Suggestive hypotheses should not be put directly to drudgery but should be entertained for a while, as rare and welcome guests.'

Weak theories usually imply the absence of precise, trustworthy data, and it is interesting that in the history of science most strong theories started out, in fact, as weak theories; their very existence stimulated the accumulation of precise data which later transformed these theories into strong ones. The heliocentric theory of Copernicus was based on wretchedly poor and inaccurate data; indeed, available observations were so erroneous and unreliable that they could not be used to arbitrate between the Copernican and the Ptolemaic theories. It was a weak theory in every respect, but by its very existence it encouraged astronomers to seek for ever more accurate observations, until Tycho Brahe finally gathered data reliable enough for Kepler to verify his three laws of planetary motion; these in turn mediated Newton's great synthesis.

This would appear to be the answer to those who feel that the publication of weak theories is premature, and should await more precise measurement and other advances in knowledge: such precise measurement, and such other advances, are usually the consequence of the interest aroused by a challenging new theory, however weak. Without the existence of the theory, there would be little motivation for the more precise measurements to be made. Every strong theory started out as a weak one (but of course not every weak theory necessarily becomes a strong one!); premature demands on such a theory for accuracy and rigour appropriate to strong theories are less likely to lead to advances than a realistic insight into the limitations of weak theories.

Roby goes on to point out that absolutist notions of validity in scientific theory find no counterpart in physics. He distinguishes three major types of validity. 'First, a theory may be valid in a *subjunctive* sense. That is, it may describe the behaviour of entities under conditions that never obtain in fact. The Galilean law of falling bodies and the ideal gas laws are examples. Second, a theory may be *locally* true, that is, may hold over certain ranges of the relevant variables. Hooke's law of elasticity is a clear-cut example, and the Newtonian laws of motion (holding for "middle-sized" phenomena) are now accepted in this sense. Finally, a law may hold *statistically*, being supported by large numbers of observations although there are local exceptions. The standard example is the law of increasing entropy. In each case, the postulated "law" can be definied as the potential limit of observational approximations as certain stipulations are satisfied.' Roby goes on to point out that 'it may well be that behavioural laws must invoke all three of these forms of licence in relating predictions to the world of observation. In other words, we may be reduced to predicting relationships among parameters of statistical distributions which can be empirically sampled within certain limits only, and which are subject to extraneous disturbing factors.'

This discussion has been rather general, and it may be of some interest to relate it more directly to the kind of problem raised by the set of theories tested in this book. We may begin by considering the order of size of expected correlations, bearing in mind the 'extraneous disturbing factors'. According to our theory, different measures of conditioning should intercorrelate positively because speed of formation of excitatory and inhibitory potential, which are fundamental to speed of conditioning, are supposed to be permanent properties of individual organisms. It might seem at first sight that a correlation of unity should be found between any two measures of conditioning. This, however, is not so. In the first place, we have the fundamental problem of deciding whether a given procedure is in fact a measure of conditioning or not. Learning theory (Cohen and Nagel's 'K') does not give us a clear answer, and the chapter by Willett in this book shows clearly the difficulties which are placed in the way of deciding whether a given procedure is in fact a measure of conditioning, or whether the observed changes are simple artefacts, or what precisely is taking place. In the second place, the classical procedures, even if we could be certain that they may justifiably be regarded as measures of conditioning, do not show very high reliabilities; figures of 0·6 or thereabouts may perhaps be regarded as typical. This would lead us to expect correlations between two such measures to be in the neighbourhood of $0{\cdot}6 \times 0{\cdot}6 = 0{\cdot}36$, which is a much lower value than unity.



A_1 = PGR Conditioning
A_2 = Introversion
B_1 = Eye-Blink Conditioning
$B_2 = B_1$
C_1 = Number of Sweat Glands
C_2 = Environmental Factors
$D_1 = D_2$ = Other Extraneous Factors
$E_1 = E_2$ = Sensitivity of Cornea
$F_1 = F_2$ = Other Extraneous Factors

But this is not all. Let A and B in the figure stand for the total non-chance variance of PGR conditioning and eye-blink conditioning scores respectively of a group of subjects. The overlap between these circles indicates the existence of a correlation (corrected for attenuation) of 0·707. The correlation fails to be unity because both sets of scores are in part determined by extraneous causes which are irrelevant to the original hypothesis, and which are specific to each set of measurements. As examples, I have given the number of sweat glands in the palm (for the PGR), and the sensitivity of the cornea (for eye-blink conditioning). D and F schematically indicate the presence of other, unidentified factors which in part determine score on these two tests of 'conditionability'.

In addition to the common, underlying factor of 'conditionability', then, we also have forces making for specificity in score on the usual type of conditioning test. These forces may be so strong as to reduce the communality to a very small figure; sometimes of course they may be quite negligible. Without knowledge of all these irrelevant extraneous causes (which should form part of 'K') it is difficult to make any precise prediction, or to interpret experimental results, particularly when these are indicative of only small or virtually non-existent relationships. It should of course be possible to measure separately the causes shown as C and E, and to treat them as suppressor variables in an equation; this would correct the observed correlation for attenuation due to this particular extraneous variable. Alternatively, it should be possible to rule out some of these variables experimentally; thus by adjusting the force of the UCS (the puff of air) in the eye-blink conditioning experiment in terms of individual thresholds and j.n.d. values for different subjects, we should be able to control this particular disturbing influence. (This obvious precaution has not been taken in the published work of Spence, Taylor, or Franks, all of whom have been concerned with the measurement of individual differences in eye-blink conditioning.)

It is more difficult to know what to do with D and F, as these letters represent hypothetical variables which are unknown and not specified by learning theory; indeed, they may be non-existent. Only the closest and most careful study of the particular conditioning measures used in the experiment will gradually elucidate all these complexities, mercifully hidden in the diagram under innocuous letters and cross-hatched figures!

The discussion will make it clear why the fact that empirically observed correlations between different conditioning measures have usually been low cannot be interpreted as being a fatal objection to the theory of the existence of a general factor of 'conditionability'. The fact that they are usually positive should be taken as a spur to purify the measures concerned, increase their reliability, and elucidate the true relationships existing between the underlying variables. Negative results at such an early stage of development are no more conclusive than were the repeated failures for several hundred years to observe stellar parallax.

When we turn to the relationship between personality variables like extraversion, and experimental variables like conditionability, similar considerations apply. The restrictions imposed by lack of reliability are just as important here as in the case of correlating one measure of conditioning with another (Eysenck, 1959). As regards the influence of extraneous variables, we may also use the same arguments as before. Let us turn back to the figure, letting A_2 represent introversion this time, and B_2 eye-blink conditioning. The only change we have to make is to replace C_1 by C_2, which we now take to represent differential environmental influences. Introversion, according to the theory, represents the effect of superior conditionability *under conditions of equal external (social) pressure*. Clearly such equality implies validity of the theory in what Roby has called a *subjunctive* sense; 'that is, it may describe the behaviour of entities under conditions that never obtain in fact'. We cannot experimentally do much to produce equality, but we can possibly use an estimate or a measure of inequality as a suppressor variable in our equations. This is being done at the moment in a research being carried out under a grant from the Home Office. The purpose of this work is to test the theory that low conditionability (extraversion) predetermines children to grow up into less strongly socialized adults, the hypothesis being that the process of socialization is mediated by conditioning. Now clearly the deduction that criminals should be more extraverted than matched non-criminals would only follow if the amount of conditioning received in the direction of socialization had been equal for all the subjects of the experiment. An introvert, brought up and conditioned in Fagin's kitchen, would presumably be more likely to become a criminal than an extravert brought up under severe disciplinary conditions in an honest home. Thus unless separate estimates are available of conditions of upbringing, the correlation between criminality and conditionability, or between criminality and extraversion, could take on almost any value at all, depending on the selection of cases.

The same consideration applies to the problem of the relation between extraversion-introversion, on the one hand, and hysteria-dysthymia, of the other. I have followed Jung in hypothethizing that extraverted personality predisposes a person to develop hysterical and psychopathic symptoms, while introverted personality predisposes a person to develop dysthymic symptoms. Such a hypothesis does not postulate a one-to-one correspondence, however; it preserves the distinction between personality, on the one hand, and symptomatology on the other. Foulds (1958) has argued, quite correctly, that the two should not be identified; this is implicit in my statement of the theory. Anxiety is regarded as a conditioned fear response, and the hypothesis linking anxiety with introversion was elaborated on the basis of postulating greater conditionability on the part of introverted people, leading to the ready elaboration of conditioned fear responses. But the development of conditioned responses is determined by other factors than the degree of conditionability of an organism; frequency of conditioning trials, stimulus-response asynchronism, delay in reinforcement, and the nature and strength of the reinforcing agent are other variables which may be vastly more important in any particular situation. By pleading for the recognition of this one factor, I did not intend to plead for the neglect of all others! Even extraverts do form conditioned responses, after all; it is only *ceteris paribus* that they condition less quickly than introverts! Except under experimental conditions (and not always then!) *cetera* are not usually *paria*, and extraverts may therefore form conditioned fear responses, while introverts fail to do so. Again, therefore, a realistic appraisal must include not only conditionability, but also an assessment of the external conditions responsible for the development of the neurotic disorder in question. It is only when this assessment is used as a suppressor variable in the correlation that we may say that introversion = dysthymia, and extraversion = hysteria/psychopathy. It would seem to follow from our theory that greater uncertainty should attach to the diagnostic classification of neurotics of extraverted personality, as these may develop anxiety and other dysthymic reactions under appropriate conditions; it is more difficult to imagine circumstances under which neurotic introverts would develop hysterical or psychopathic symptoms. Foulds' results confirm this prediction. It is to be hoped that his pioneer experiment will be followed up by others, in an attempt to test this general hypothesis; the relation between personality and symptom development has been curiously neglected by psychiatrists and psychologists alike.

In summary of this discussion, it might perhaps be fair to repeat that when we are dealing with a strong theory, successful predictions are commonplace and do not do much to enhance the validity of the theory, while failures to predict correctly are very serious and may be disastrous for the theory in question. When we are dealing with a weak theory, however, the reverse is true. Successful predictions, particularly if unlikely on commonsense grounds, are important and valuable, and do much to support the right of the theory to be considered seriously. Failures, however, are not unequivocal enough to be taken too seriously; they certainly do not damage the theory in the same way as do failures in the case of a strong theory.

We may put this point of view in terms of information theory. If we may take K as certain (or near certain), then the confirmation or infirmation of H gives us one bit of information. (Provided both outcomes are equiprobable.) If we take K as quite uncertain, then the failure of H to be verified may have many causes; its success, however, is unlikely to have a cause other than the one specified. Consequently, success gives us more information than failure. It gives us information, not only about H, but also about K. Success implies that both H and K are essentially correct; failure implies that either H, or any part of K may be in error, and this is too vague to be very useful. Failure may of course be used to suggest more complex relationships, improvements in technique, or limitations in choice of parameters, but it is not in itself very informative. Information gained by successful tests of a strong theory is largely redundant, while failure is highly informative. The reverse is true of weak theories.[1]

---

[1] An example from physics may make this statement clearer. There are certain stars, the so-called 'white dwarfs', to which Eddington's mass-luminosity principle does not apply; they are supposed to be much more massive than might be expected from their luminosity. (This is because in them matter reaches an extraordinary high degree of density at which the ordinary gas laws cease to apply.) Use has been made of this condition to furnish another proof for Relativity theory. This theory predicts that the apparent frequency of a periodic phenomenon—such as atomic vibration—is changed when the source of the vibration is situated in a strong gravitational field. The

What has been said in this chapter should not of course be taken to mean that the psychologist has unbridled licence to theorize to his heart's content, regardless of negative results. For a theory to be regarded seriously it must show some evidence, both internally and externally, that it can mediate experimental predictions. But once this evidence is available, early failures of *some* predictions should not be taken too seriously; they should be regarded as a challenge to discover the causes of the failure, rather than as necessarily disproving the theory. This, of course, implies a tolerance of ambiguity, an ability to suspend judgement, which is often difficult to maintain. Nevertheless, the whole history of science argues powerfully against rash decisions on the basis of inadequate data; time and time again we have seen buried theories rise afresh from the ashes to which they had been prematurely consigned.

Nor should expectations be too high of close relationships between experimental variables. Much refinement will have to go into the raw scores obtained from psychological tests before we can begin to claim to measure one single variable, rather than a mixture of the most diversified traits, abilities and attitudes. The best that can be expected is a set of low correlations, usually in the expected direction, but occasionally directly opposed to prediction; on such a foundation we can then begin to erect the infinitely complex set of laws and functional relationships, concepts and definitions, which will ultimately, shorn of ambivalences and ambiguities, constitute that proper science of behaviour and personality which so obviously does not exist at the present time, except possibly as a foundation for a palimpsest.

---

slowing down of the vibration shows itself in a slight displacement of spectral lines towards the red; it was observed in 1925 by Adams in the spectrum of the Companion of Sirius. Vaucouleurs (1957) points out: 'As the observed effect was in good agreement with the theoretical prediction, both the theory and the existence of extremely dense matter in white dwarfs were confirmed.' Here the relativity theory (H_1) is being tested in relation to a body of knowledge (K) not itself too firmly established; confirmation of the hypothesis is likely only if both H_1 and K are correct, and consequently contributes much to our knowledge of both relativity theory and of the nature of white dwarfs. Failure of the hypothesis could have been due to errors in the relativity theory, or to lack of true knowledge of the composition of white dwarfs; it would therefore been ambiguous at best, and not very informative.

BIBLIOGRAPHY AND AUTHOR INDEX

Numbers in italic type at the end of each entry refer to the page on which the reference is quoted.

AITKEN, A. C. *Determinants and Matrices.* London: Oliver & Boyd, 1939. *299.*

AMMONS, R. B., & AMMONS, C. H. Bilateral transfer of rotary pursuit skill. *Amer. Psychologist*, 1951, 6, 294. *109.*

AX, A. F., & COLLEY, W. H. Temporal acuity of vision, audition, and touch in psychogenic and neurogenic pathology. *J. consult. Psychol.*, 1955, 19, 455–61. *227.*

BABCOCK, HARRIET. An experiment in the measurement of mental deterioration. *Archives of Psychology*, 1930, 18, No. 117. *5, 6, 9, 10, 12, 19, 23, 68.*

BABCOCK, HARRIET. *Dementia Praecox: A Psychological Study.* The Science Press, 1933. *5, 6, 9, 10, 12, 19, 23, 68.*

BABCOCK, HARRIET. *Time and the Mind.* Cambridge, Mass.: Sci-Art Publishers, 1941. *5, 6, 9, 10, 12, 19, 23, 68.*

BABCOCK, HARRIET & LEVY, LYDIA. *Manual of Directions for the Revised Examination of the Measurement of Efficiency of Mental Functioning.* Chicago: Stoelting, 1940. *5, 6.*

BAKAN, P. Extraversion-introversion and improvement in an auditory vigilance task. *Brit. J. Psychol.* 1960, 50, 325–32. *114.*

BARLOW, H. B. Eye movements during fixation. *J. Physiol.*, 1952, 116, 290–306. *212.*

BARTLETT, M. S. Tests of significance in factor analysis. *Brit. J. Stats. Psych.*, 1950, 3, 77–84. *274.*

BECKER, W. C. A genetic approach to the interpretation and evaluation of the process-reactive distinction in schizophrenia. *J. abnorm. soc. Psychol.*, 1956, 53, 229–36. *7.*

BEECH, H. R. *An Investigation of the Performance on a Perceptual-motor Task by Psychiatric Patients with Special Reference to Brain Damage.* Unpublished Ph.D. Thesis, 1956. *123.* University of London Library. *62.*

BEKESY, G. VON, & ROSENBLITH, W. A. The mechanical properties of the ear. In S. S. Stevens (ed.). *Handbook of experimental Psychology.* New York: Wiley, 1951. Pp. 1075–115. 228.

BELSON, W. A. *The Construction of an Index of Intelligence.* London: 1955, B.B.C. Audience Research Dept. *17.*

BENDER, M. B., & TEUBER, H. L. Disorders in visual perception following cerebral lesions. *J. Psychol.*, 1949, 28, 223–33. *205.*

BENDER, M. B., & TEUBER, H. L. Spatial organisation of visual perception following injury to the brain. *Arch. Neurol. Psychiat.*, 1947, 58, 721–39. 1948, 59, 39–62. *205.*

BENJAMIN, J. D. A method for distinguishing and evaluating formal thinking disorders in schizophrenia. In J. S. Kasanin (ed.) *Language and Thought in Schizophrenia.* Berkeley and Los Angeles: Univ. of California Press, 1946. *7, 10, 28.*

BERGMAN, P. S., NATHANSON, M., & BENDER, M. B. Electrical recordings of normal and abnormal eye movements modified by drugs. *Arch. Neurol. Psychiat.*, 1952, 67, 357–74. *212.*

BEVAN, W., jr. The influence of figural after-effects upon visual intensity thresholds. *J. gen. Psychol.*, 1951, 45, 189–207. *221.*

BILLS, A. G. Blocking: a new principle of mental fatigue. *Amer. J. Psychol.*, 1931, 43, 230–45. *114.*

BINDRA, D., PATERSON, A. L., & STRZELECKI, J. On the Relationship between anxiety and Conditioning. *Canadian J. Psychol.*, 1955, 8, 1–6. *167.*

BITTERMAN, M. E. A reply to Dr. Luckiesh. *J. exp. Psychol.*, 1936, 46, 182–4. *212.*

BITTERMAN, M. E., & SOLOWAY, E. Frequency of blinking as a measure of visual efficiency: some methodological considerations. *Amer. J. Psychol.*, 1946, 55, 676–81. *212.*

BLEULER, E. Dementia praecox. New York: Int. Univ. Press, 1950. *3.*

BOLLES, MARJORIE, & GOLDSTEIN, K. A study of impairment of 'abstract behaviour' in schizophrenic patients. *Psychiat. Quart.*, 1938, 12, 42–65. *6.*

BRENGELMANN, J. C. Der visuelle Objecterkennungstest. *Z. f. Exp. u. angew. Psychol.*, 1, 422–52. *132, 146.*

BRENGELMANN, J. C. Validierung des Bilderkennung im Gruppenversuch. *Z. f. Diagnostiche Psychol. u. Personlich Keitsforschung*, 6, 3–17. *132, 146.*

BRENNER, MAY W. Continuous stimulation and apparent movement. *J. Psychol.*, 1953, 66, 494–5. *217.*

BROADBENT, D. E. *Perception and Communication.* London: Pergamon Press, 1958. *113, 114, 136.*

BROADHURST, ANN, & BROADHURST, P. L. An analysis of the pursuit rotor learning of psychotics under massed and spaced practice. Paper read to *Two-day Conference of the B.P.S.*, London (December, 1958). *138, 139.*

BROMLEY, D. B. Some experimental tests of the effect of age on creative intellectual output. *J. Gerontol.*, 1956, 11, 74–82. *42.*

BRUELL, J. H., & ALBEE, G. W. Notes toward a motor theory of visual egocentric localisation. *Psychol., Rev.*, 1955, 62, 391–400. *212.*

BRUNER, J. S., GOODENOW, J. J., & AUSTIN, G. A. A Study of Thinking. John Wiley & Sons, Inc., New York, 1956. *175, 181.*

CAMERON, N. Reasoning, regression and communication in schizophrenics. *Psychol. Monogr.*, 1938a, 50, No. 221, 1–33. *7, 8.*

CAMERON, N. A study of thinking in senile deterioration and schizophrenic disorganization. *Amer. J. Psychol.*, 1938b, 51, 650–64. *7, 8.*

CAMERON, N. Deterioration and regression in schizophrenic thinking. *J. abnorm. soc. Psychol.*, 1939a, 34, 265–70. *7, 8.*

CAMERON, N. Schizophrenic thinking in a problem-solving situation. *J. ment. Sci.*, 1939b, 85, 1012–35. *7, 8.*

CAMERON, N. *The Functional Psychoses.* In J. McV. Hunt (ed.) Personality and the Behavior Disorders. Vol. II. New York, 1944: Ronald Press Co. *7, 8.*

CAMERON, N. *The Psychology of Behavior Disorders.* Boston: Houghton Mifflin, 1947. *7, 8.*

CAMERON, N., & MARGARET, ANN. Experimental studies in thinking: I. Scattered speech in the responses of normal subjects to incomplete sentences. *J. exp. Psychol.*, 1949, 39, 617–27. *7, 8.*

CAMERON, N., & MARGARET, ANN. Correlates of scattered speech in the responses of normal subjects to incomplete sentences. *J. gen. Psychol.*, 1950, 77–84. *7, 8.*

CAMERON, N., & MARGARET, ANN. *Behavior Pathology.* Boston: Houghton Mifflin, 1951. *7, 8.*

CARLSON, V. R. Satiation in a reversible perspective figure. *J. exp. Psychol.*, 1953, 45, 442–8. *217.*

CATTELL, R. B., KING, J. E., & SCHUETTLER, A. R. Contact Personality Questionnaire. Champaign, Illinois: Institute for Personality and Ability Testing, 1954. *198,199.*

CHAPMAN, L. F., SYMMES, D., & HALSTEAD, W. C. Flutter fusion in patients with cortical ablations. *J. comp. Physiol., Psychol.*, 1955, 48, 421–5. *227.*

CHAPMAN, L. J. Distractibility in the conceptual performance of schizophrenics. *J. abnorm. soc. Psychol.*, 1956, 53, 286–91. *8.*

CHAPMAN, L. J., & TAYLOR, JANET A. Breadth of deviate concepts used by schizophrenics. *J. abnorm. soc. Psychol.*, 1957, 54, 118–23. *8.*

CHRISTMAN, R. J. Shifts in pitch as a function of prolonged stimulation with pure tones. *Amer. J. Psychol.*, 1954, 67, 484–7. *217.*

COHEN, M. R., & NAGEL, N. *An Introduction to Logic and Scientific Method.* New York: Harcourt, Brace & Co., 1936. *307.*

COMREY, A. L. A factor analysis of items on the K scale of the M.M.P.I. *Educ. Psychol. Measmt.* In press. *199.*

CRUTCHFIELD, R. S., & EDWARDS, W. The effect of a fixated figure on autokinetic movement. *J. exp. Psychol.*, 1949, 39, 561–8. *217.*

DAS, J. P. An Experimental Study of the Relation between Hypnosis Conditioning and Reactive Inhibition. *Thesis submitted for Ph.D. Univ. London, 1957. 165.*

DASTON, P. G., KING, G. F., & ARMITAGE, S. G. Distortion in paranoid schizophrenia. *J. consult. Psychol.*, 1953, 17, 50–3. *160.*

DEATHERAGE, B. H., & BITTERMAN, M. E. The effect of satiation on stroboscopic movement. *Amer. J. Psychol.*, 1952, 65, 108–9. *217.*

DEUTSCH, J. A. A new phenomenon of apparent movement and its after-effect. *Quart. J. exp. Psychol.*, 1950, 2, 119–23. *203, 204.*

DEUTSCH, J. A. A preliminary report on a new auditory after-effect. *Quart. J. exp. Psychol.*, 1951, 3, 43–6. *217.*

DITCHBURN, R. W., & GINSBERG, B. L. Vision with a stabilized retinal image. *Nature*, 1952, 170, 36–7. *212.*

DITCHBURN, R. W., & GINSBERG, B. L. Involuntary eye movements during fixation. *J. Physiol.*, 1953, 119, 1–17. *212.*

DUNCAN, C. P. On the similarity between reactive inhibition and neural satiation. *Amer. J. Psychol.*, 1956, 69, 227–35. *122, 194.*

EPSTEIN, S. Overinclusive thinking in a schizophrenic and a control group. *J. consult. Psychol.*, 1953, 17, 384–8. *7, 18, 51.*

EYSENCK, H. J. *Dimensions of Personality.* London: Kegan Paul, 1947. *133, 275.*

EYSENCK, H. J. Schizothymia-cyclothymia as a dimension of personality: I. Historical review. II. Experimental. *J. Personality*, 1952a, 19, 123–52; 20, 345–84. *11, 68, 71.*

EYSENCK, H. J. *The Scientific Study of Personality.* London: Routledge & Kegan Paul, 1952b. *11.*

EYSENCK, H. J. The logical basis of factor analysis. *Amer. Psychologist*, 1953a, 8, 105–114. *275.*

EYSENCK, H. J. *The structure of human personality.* London, Methuen, 1953b. *112, 133, 193, 194.*

EYSENCK, H. J. La rapidite du functionnement mental comme mesure de l'anomalie mentale. *Rev. de Psychol. appl.*, 1953c, 13, 367–77. *66.*

EYSENCK, H. J. Cortical inhibition, figural after-effect, and theory of personality. *J. abnorm. soc. Psychol.*, 1955a, 51, 94–106. *108, 123, 194, 217.*

EYSENCK, H. J. A dynamic theory of anxiety and hysteria. *J. ment. Sci.*, 1955b, 101, 28–51. *66, 108, 204.*

EYSENCK, H. J. Psychiatric diagnosis as a psychological and statistical problem. *Psycho. Reports*, 1955c, 1, 3–17. *204.*

EYSENCK, H. J. The questionnaire measurement of neuroticism and extraversion. *Rev. Psychol.*, 1956a, 54, 113–40. *108, 198, 205.*

EYSENCK, H. J. Reminiscence, drive, and personality theory. *J. Abnorm. soc. Psychol.*, 1956b, 53, 328–33. *119, 136, 139.*

EYSENCK, H. J. 'Warm-up' in pursuit rotor learning as a function of the extenction of conditioned inhibition. *Acta. Psychol.*, 1956c, 12, 349–70. *139, 141.*

EYSENCK, H. J. *The dynamics of anxiety and hysteria.* London: Routledge & Kegan Paul, 1957. *108, 122, 123, 157, 211, 271.*

EYSENCK, H. J. A short questionnaire for the measurement of two dimensions of personality. *J. Appl. Psychol.*, 1958a, 42, 14-17.

EYSENCK, H. J. Hysterics and dystymics as criterion groups in the measure of introversion-extraversion: A reply. *J. abnorm. soc. Psychol.*, 1958b, 57, 250-252. *112.*

EYSENCK, H. J. Comments on a test of the personality-satiation-inhibition theory. *Psychol. Rep.*, 1959a, 5, 395-6. *239.*

EYSENCK, H. J. *Manual of the Maudsley Personality Inventory.* London: Univ. of London Press, 1959b. *159, 198, 242.*

EYSENCK, H. J. Serial position effects in nonsense syllable learning as a function of interlist rest pauses. *Brit. J. Psychol.*, 1959c, 50, 360-2 *306.*

EYSENCK, H. J. *The Structure of Human Personality* (2nd ed.), London: Methuen, 1960. *243.*

EYSENCK, H. J., GRANGER, G. W., & BRENGELMANN, J. C. *Perceptual Processes and Mental Illness.* London: Maudsley Monographs, No. 2. Chapman & Hall, 1957. *11, 66, 146, 193, 194.*

EYSENCK, H. J., HOLLAND, H. C., & TROUTON, D. S. Drugs and personality. III. The effects of stimulant and depressant drugs on visual after-effects. *J. ment. Sci.*, 1957, 103, 650-55. *203, 204.*

EYSENCK, H. J., & HOLLAND, H. Two measures of figural after-effect. *Indian J. Psychol.*, 1958, 33II, 85-92. *218.*

EYSENCK, H. J., & SLATER, P. Effects of practice and rest on fluctuations in the Muller-Lyer illusion. *Brit. J. Psychol.*, 1958, 49, 246-56. *217.*

EYSENCK, SYBIL, B. G. *A Dimensional Analysis of Mental Abnormality.* Unpublished Ph.D. Thesis, 1955, Univ. of London Library. *12, 66, 68, 69, 70, 71, 72, 73, 76, 77.*

FARBER, J. E., & SPENCE, K. W. Complex Learning and Conditioning as a function of Anxiety. *J. Experiment. Psychol.*, 1953, 45, 120-5. *161.*

FELDMAN, M. J., & Drasgow, J. A visual-verbal test for schizophrenia. *Psychiat. Quart. Suppl.*, Part 1, 1951, 1-10. *10.*

FEY, ELIZABETH T. The performance of young schizophrenics and young normals on the Wisconsin card sorting test. *J. consult. Psychol.*, 1951, 15, 311-19. *6.*

FINESINGER, J. E., SUTHERLAND, G. F., & MCGUIRE, F. F. The Positive Conditional Salivary Reflex in Psychoneurotic Patients. *Amer. J. Psychiat.*, 1942, 99, 61. *167.*

FISHER, R. A. Statistical Methods for Research Workers. *Oliver & Boyd. Edinburgh*, 1948. *161.*

FISHER, S. Patterns of personality rigidity and some of their determinants. *Psychol. Monogr.*, 1950, 64, No. 1 (whole No. 307). *6.*

FOULDS, G. A. Temperamental differences in maze performance, Part I, Characteristic differences among psychoneurotics. *Brit. J. gen. Psych.*, 1951, 42, 209-217. *298.*

FOULDS, G. A. A method of scoring the T.A.T. applied to psychoneurotics. *J. ment. Sci.*, 1953, 99, 235-46. *298.*

FOULDS, G. A., & CAINE, T. M. Psychoneurotic symptom clusters, tait clusters and *psychological tests. J. ment. Sci.*, 1958, 104, 722-31. *313.*

FOX, L. Practical methods for the solution of linear equations and the inversion of matrices. *J. Roy. Stats. Soc.*, 1950, 12b, 120-48. *299.*

FRANKS, C. M. An Experimental Study of Conditioning as related to mental abnormality. Ph.d. Thesis, Univ. London, 1954. *160, 165, 166, 173.*

FRANKS, C. M. Conditioning and Personality: A Study of normal and neurotic Subjects. *J. Abn. & Soc. Psychol.*, 1956, 52, 143-50. *108, 157, 160, 161.*

FRANKS, C. M. Personality Factors and rate of conditioning. *Brit. J. Psychol.*, 1957, 48, 119-26. *160, 165.*

FRANKS, C. M. Effect of food, drink and Tobacco deprivation on the conditioning of the eyeblink response. *J. Exper. Psychol.*, 1957a, 53, 117-20. *157, 160.*

FRANKS, C. M. Personality Factors and rate of conditioning. *Brit. J. Psychol.*, 1957b, 48, 119-26. *15, 160.*

FRANKS, C. M., & LAVERTY, S. G. Sodium amytal and eyelid Conditioning. *J. Ment. Sci.* 101, 1955, 654–63. *160.*

FRANKS, C. M., SOUIEF, M. J., & MAXWELL, A. M. A Factorial Study of Certain Scales from the M.M.P.I. and the S.T.D.C.R. (to appear). *160, 180, 191.*

FRANKS, C. M., & TROUTON, D. Effects of Amylobabital Sodium and Dexamphetamine Sulphate on the conditioning of the eyeblink response. *J. Comp. Physiol. Psychol.*, 1958, 51, 220–2. *160.*

FRANKS, C. M., & WILLETT, R. A. The Interruption and Resumption of an Eyeblink Conditioning Schedule. 1959, unpublished manuscript. *164.*

FRANKS, C. M., & WITHERS, W. C. R. Photoelectric recording of Eyelid Movements. *Amer. J. Psychol.*, 1955, 68, 467–71. *161.*

FREEBURNE, C. M., & HAMILTON, C. F. The effect of brightness on figural after-effects. *Amer. J. Psychol.*, 1949, 62, 567–9. *217.*

FREEMAN, E., & JOSEY, W. E. Quantitative visual index to memory impairment. *Arch. Neurol. Psychiat.*, 1949, 62, 794–7. *205.*

FREUD, S. *A general introduction to psychoanalysis*. New York: Liveoright, 1920. *271*

FRUCHTER, B. *Introduction to factor analysis*. New York: D. Van Nostrand, 1954. *254.*

FURNEAUX, W. D. *Manual of Nufferno Speed Tests*. London: Nat. Foundation of Educational Research, 1956a. *5, 9, 19.*

FURNEAUX, W. D. *Manual of Nufferno Level Tests*. London: Nat. Foundation of Educational Research, 1956b. *5, 9, 19.*

GALLESE, G. J. Spiral after-effect as a test of organic brain damage. *J. clin. Psychol.*, 1956, 12, 254–8. *123, 205.*

GANTT, W. H. The nervous secretion of Saliva. Quantitative studies in the natural unconditioned reflex secretion of saliva. *Amer. J. Physiol.*, 1936, 116, 56–7. *167.*

GANTT, W. H. The nervous secretion of saliva. The relation of the conditioned reflex to the intensity of the unconditioned stimulus. *Amer. J. Physiol.*, 1938, 123, 74. *167.*

GATES, L. W. The after-effect of visually observed movement. *Amer. J. Psychol.*, 1934, 46, 34–46. *203.*

GEBHARD, J. W., MOWBRAY, G. H., & BYHAM, C. L. Difference limens for photic intermittance. *Quart. J. exp. Psychol.*, 1955, 7, 49–55. *227.*

GEORGE, F. H. On the theory of the figural after-effects *Canad. J. Psychol.*, 1953, 7, 167–71. *216.*

GEORGE, F. H. On the figural after-effect. *Quart. J. exp. Psychol.*, 1953, 5, 128–35. *216.*

GIBSON, J. J. Adaptation, after-effect, and contrast in the perception of tilted lines. II. Simultaneous contrast and the areal restriction of the after-effect. *J. exp. Psychol.*, 1937, 20, 553–68. *216.*

GIBSON, J. J. Adaptation with negative after-effect. *Psychol., Rev.*, 1937, 44, 222–44. *216.*

GIBSON, J. J. The visual perception of objective motion and subjective movement. *Psychol., Rev.*, 1954, 61, 304–14. *216.*

GIBSON, J. J., & WADDELL, D. Homogeneous retinal stimulation and visual perception. *Amer. J. Psychol.*, 1952, 65, 263–70. *216.*

GOLDBERG, L. R., & SMITH, P. A. The clinical usefulness of the Archimedes spiral in the diagnosis of organic brain damage. *J. consult. Psychol.*, 1958, 22, 153–7. *205.*

GOLDSTEIN, K. The significance of special mental tests for diagnosis and prognosis in schizophrenia. *Amer. J. Psychiat.*, 1939, 96, 575–8. *6, 7, 10, 26, 33.*

GOLDSTEIN, K. Methodological approach to the study of schizophrenic thought disorder. In J. S. Kasanin (ed.), *Language and Thought in Schizophrenia*. Berkeley and Los Angeles: Univ. of California Press, 1946. *6, 7, 10, 26, 33.*

GOLDSTEIN, K., & SCHEERER, M. Abstract and concrete behaviour, an experimental study with special tests. *Psychol. Monogr.*, 1941, 53, No. 2. *6, 7, 10, 26, 33, 134.*

GOLDSTONE, S. Flicker fusion measurements and anxiety level. *J. exp. Psychol.*, 1955, 49, 200–2. *242.*

GOODFELLOW, L. D. The stability of auditory and vibro-tactile thresholds. *J. gen. Psychol.*, 1938, 18, 49–55. *228.*

GORHAM, D. R. Use of the proverbs test for differentiating schizophrenics from normals. *J. consult. Psychol.*, 1956, 20, 435–40. *7.*

GRANGER, G. W. Personality and visual perception: a review. *J. ment. Sci.*, 1953, 99, 8–43. *200, 241.*

GRANGER, G. W. The night visual ability of psychiatric patients. *Brit. J. physiol. Opt.*, 1954, 11, 238–44. *241.*

GRANGER, G. W. Dark adaptation in neurotic patients. *J. ment. Sci.*, 1955, 101, 354–62. *241.*

GRANGER, G. W. Night vision and psychiatric disorder: a review of experimental studies. *J. ment. Sci.*, 1957, 103, 48–79. *241.*

GRANIT, R. On inhibition in the after-effect of seen movement. *Brit. J. Psychol.*, 1928, 19, 147. *203, 204.*

GRANT, D. A., HORNSETH, J. P., & HAKE, H. W. The influence of the inter-trial interval on the Humphrey's random reinforcement effect during the extinction of a verbal response. *J. Exper. Psychol.*, 1950. 40, 609–12. *174, 176.*

GRAVELY, A. M. Unpublished Doctor's dissertation. University of London, 1950. *201.*

GRINDLEY, G. C. The relation between the rod and cone mechanisms in the after-effect of seen movement. *J. Physiol.*, 1930, 69, 53–59. *203, 204.*

GRINDLEY, G. C., & WILKINSON, R. T. The after-effect of seen movement on a plain field. *Quart. J. exp. Psychol.*, 1953, 5, 183–184. *204.*

GUILFORD, J. P., & PARK, D. G. The effect of interpolated weights upon comparative judgements. *Amer. J. Psychol.*, 1931, 43, 589–99. *125.*

GUTHRIE, E. R. The Psychology of Learning. *Rev. Ed.*, 1952, *Harper Bros., N.Y.* 189.

GWYNNE-JONES, H. G. The status of inhibition in Hull's system: a theoretical revision. *Psychol. Rev.*, 1958, 65, 179–182. *140.*

HAMMER, E. R. Temporal factors in figural after-effects. *Amer. J. Psychol.*, 1949, 62, 337–54. *216.*

HANFMANN, EUGENIA, & KASANIN, J. A method for the study of concept formation. *J. Psychol.*, 1937, 3, 521–40. *6.*

HANFMANN, EUGENIA, & KASANIN, J. Conceptual thinking in schizophrenia. *Nerv. Ment. Dis. Monogr.*, 1942, 67. *6.*

HATHAWAY, S. R., & MCKINLEY, J. C. *The Minnesota Multiphasic Personality Inventory.* New York: The Psychological Corpn., 1943. *199.*

HATHAWAY, S. R., & MCKINLEY, J. C. Minnesota Multiphasic Personality Inventory Manual revised 1951. *159.*

HEBB, D. O. *The Organisation of Behaviour: a neuro-psychological theory.* New York: Wiley, 1949. *136, 137, 218.*

HEBB, D. O. Drives and the CNS (Conceptual Nervous System). *Psychol. Rev.*, 1955, 62, 243–54. *136, 137.*

HERON, W., DOANE, B. K., & SCOTT, T. H. Visual disturbance after prolonged perceptual isolation. *Canad. J. Psychol.*, 1956, 10, 13–18. *200, 211.*

HILDEBRAND, H. P. A factorial study of introversion-extraversion by means of objective tests. *Ph.D. Thesis*, Univ. London, 1953. *271.*

HILDEBRAND, H. P. A factorial study of introversion-extraversion. *Brit. J. gen. Psychol.*, 1958, 49, 1–12. *271.*

HIMMELWEIT, H. T., DESAI, M., & PETRIE, A. An experimental investigation of neuroticism. *J. Pers.*, 1946, 15, 173. *201.*

HOCHBERG, J. E. Figure-ground reversal as a function of visual satiation. *J. exp. Psychol.*, 1950, 40, 682–6. *217.*

HOCHBERG, J. E., & TRIEBAL, W. FIGURAL after-effects with coloured stimuli. *Amer. J. Psychol.*, 1955, 68, 133–5. *221.*

HOFFMAN, A. C. Eye movements during prolonged reading. *J. exp. Psychol.*, 1946, 36, 95–118. *212.*

HOLLAND, H. C. The Archimedes spiral. *Nature*, 1957, *179*, 432–3. *147, 203, 204.*

HOLLAND, H. C. Some determinants of seen after movement in the Archimedes Spiral. *Acta Psychol.*, 1958, 14, 215–22. *124, 147, 203, 204.*

HOLLAND, H. C., & BEECH, H. R. The spiral after-effect as a test of brain damage. *J. ment. Sci.*, 1958, 104, 466–71. *204, 205, 212.*

HOLZINGER, K. J., & HARMAN, H. H. *Factor analysis.* Chicago: Univer. Chicago Press, 1941. *251.*

HOWELLS, T. H. The Experimental development of colour-tone synaesthesia. *J. Exp. Psychol.*, 1944, 99, 660–7. *183.*

HOWSON J. D. Intellectual impairment associated with brain injured patients as revealed in the Shaw test of abstract thought. *Canad. J. Psychol.*, 1948, 2, 125–33. *42.*

HULL, C. L., *et al.* Mathematico Deductive theory of rate learning—a study in Scientific methodology. *New Haven: Yale University Press, 1940. 161.*

HULL, C. *Principles of behaviour.* New York: D. Appleton-Century, 1943. *108, 109, 161, 194, 275.*

HULL, C. L. A Behaviour System. *New Haven: Yale University Press, 1952. 161.*

HUMPHREYS, L. G. Acquisition and Extinction of verbal Expectations in a situation analogous to conditioning. *J. Exper. Psychol.*, 1939, 25, 294–7. *174.*

HUNT, J. MCV., & COFER, C. N. Psychological deficit. In Hunt, J. McV. (ed.), *Personality and the Behaviour Disorders.* New York: Ronald Press, 1944. *115.*

HUNTER, W. S. The after-effect of visual motion. *Psychol. Rev.*, 1914, 21, 245–77. *203.*

HUNTER, W. S. Retinal factors in visual after-movement. *Psychol., Rev.*, 1915, 22, 479–489. *203.*

HUSTON, P. E., & SHAKOW, D. Learning in schizophrenia: I. Pursuit learning. *J. Pers.*, 1948, 17, 52–74. *115, 120.*

HUSTON, P. E., SHAKOW, D., & RIGGS, L. A. Studies of motor function in schizophrenia: II. Reaction time. *J. gen. Psychol.*, 1937, 16, 39–82. *115–120.*

IKEDA, H., & OBONAI, T. The quantitative analysis of figural after-effects: I. The process of growth and decay of figural after-effects. *Jap. J. Psychol.*, 1953, 23, 246–60. *216.*

IKEDA, H., & OBONIA, T. The quantitative analysis of figural after-effects: II. On 'self satiation'. *Jap. J. Psychol.*, 1953, 24, 179–92. *216.*

JAFFE, R. Kinesthetic after-effects following cerebral lesions. *Amer. J. Psychol.*, 1954, 67, 668–76. *217.*

JAFFE, R. The influence of visual stimulation on kinesthetic figural after-effects. *Amer. J. Psychol.*, 1956, 69, 70–75. *217.*

JENSEN, A. R. The Maudsley Personality Inventory. *Acta Psychol.*, 1958, 14, 314–25. *198, 199.*

JOHNSON, D. E., BAUER, R. W., & BROWN, D. R. Visual after phenomena in diagnosis. *J. consult. Psychol.*, 1959, 23, 90. *205.*

JONES, O. M. The relationship between visual and auditory discrimination and anxiety level. Doctor's Dissertation 1956 (Duke University). Report of A.P.A. Dept., 1956. *227.*

JONES, F. N., & BUNTING, ELIZABETH, B. Displacement after-effects in auditory localization. *Amer. Psychologist*, 1949, 4, 389. *217.*

JUNG, C. J. *Psychological types.* London: Kegan Paul, 1924. *271.*

KASANIN, J. S. The disturbance of conceptual thinking in schizophrenia. In J. S. Kasanin (ed.) *Language and Thought in Schizophrenia.* Berkeley and Los Angeles: Univ. of California Press, 1946. *6.*

KASANIN, J., & HANFMANN, EUGENIA. Disturbances in concept formation in schizophrenia. *Archives of Neurol. & Psychiat.*, 1938a, 40, 1276–82. *6.*

KASANIN, J., & HANFMANN, EUGENIA. An experimental study of concept formation in schizophrenia. *Amer. J. Psychiat.*, 1938b, 95, 35–52. *6.*

KELLY, E. LOWELL. An attempt to produce artificial chronaesthesia by the technique of the conditioned response. *J. Exp. Psychol.*, 1937, 17, 315–41. *183.*

KENDALL, M. G. *A course in multivariate analysis*. London: Charles Griffin, 1957. *259, 288, 299.*

KIMBLE, G. A. An experimental test of a two-factor theory of inhibition. *J. exp. Psychol.*, 1949, 39, 15–23. *109.*

KIMBLE, G. A. Transfer of work inhibition in motor learning. *J. exp. Psychol.*, 1952. 43, 391–2. *109.*

KIMBLE, G. A., & SHATEL, R. B. The relationship between two kinds of inhibition and the amount of practice. *J. exp. Psychol.*, 1952, 44, 355–9. *109.*

KLEIN, G. S., & KRETCH, D. Cortical conductivity in the brain injured. *J. Pers.*, 1952, 21, 118–48. *11.*

KOHLER, W. Zur Theorie des Sukzessivvergleichs und der Zeitfehler. *Psychol. Forsch.*, 1923, 4, 115–75. *123.*

KOHLER, W. Relational determination in perception. In L. A. Jeffress (ed.) *Cerebral mechanisms in behaviour*: The Hixon symposium. New York: Wiley, 1951. *217.*

KOHLER, W., & FISHBACK, JULIA. The destruction of the Muller-Lyer illusion in repeated trials. I. An examination of two theories. *J. exp. Psychol.*, 1950a, 40, 267–81. *217.*

KOHLER, W., & FISHBACK, JULIA. The destruction of the Muller-Lyer illusion in repeated trials. II. Satiation patterns and memory traces. *J. exp. Psychol.*, 1950b, 40, 398–410. *217.*

KOHLER, W., HELD, R., & O'CONNELL, D. N. An investigation of cortical currents. *Proc. Amer. phil. Soc.*, 1952. 96, 290–330. *221.*

KOHLER, W., & WALLACH, H. Figural after-effects: An investigation of visual processes. *Proc. Amer. Phil. Soc.*, 1944, 88, 269–357. *193, 194, 216, 217, 221.*

KRAUSKOPF, J. Figural after-effects in auditory space. *Amer. J. Psychol.*, 1954, 67, 278–87. *217.*

KREEZER, G. The neurological level of the factors underlying time errors. *Amer. J. Psychol.*, 1938, 51, 18–43. *123.*

KRETSCHMER, T. Körperbon und Charakter. Berlin: Springer, 1928. *3–5, 11.*

KRUGMAN, H. E. Flicker fusion frequency as a function of anxiety reaction: an exploratory study. *Psychosom. Med.*, 1947, 9, 269–72. *132, 242.*

LANCZOS, C. *Applied analysis*. London: Pitman, 1957. *299.*

LANDIS, C. Determinants of the critical flicker fusion threshold. *Physiol. Rev.*, 1954, 34, 259–86. *227.*

LANGFELD, H. S. 'Synaesthesia.' *Psychol. Bull.*, 1926, 23, 579–602. *183.*

LANGFELD, H. S. 'Synaesthesia.' *Psychol. Bull.*, 1929, 26, 582–5. *183.*

LASHLEY, K. S., CHOW, K. L., & SEMMES, J. An examination of the electrical field theory of cerebral integration. *Psychol. Rev.*, 1951, 58, 123–36. *218.*

LAWLEY, D. N. The maximum likelihood method of estimation of factor loadings, c.f. Thomson, op. cit., 1950. *274.*

LEITER, R. G., & PARTINGTON, J. *Examination Manual for the Leiter-Partington Adult Scale*. Psychological Service Center Press, 1950. *59, 60.*

LICKLIDER, J. C. R. Basic correlates of the auditory stimulus. In S. S. Stevens (ed.), *Handbook of experimental psychology*. New York: Wiley, 1951, pp. 985–1039. *228.*

LONDON, I. D. Research on Sensory Interaction in the Soviet Union. *Psychol. Bull.*, 1954, 51, 531–68. *183.*

LUCHINS, A. S. Mechanization in problem solving: the effect of Einstellung. *Psychol. Monogr.*, 1942, 54, No. 6. *8, 55, 56.*

LUCHINS, A. S., & LUCHINS, EDITH H. The satiation theory of figural after-effects and the Gestalt principles of perception. *J. gen. Psychol.*, 1953a, 49, 3–29. *18, 55, 56, 218.*

LUCHINS, A. S., & LUCHINS, EDITH H. The satiation theory of figural after-effects and the principle of pragnanz. *J. gen. Psychol.*, 1953b, 49, 185-9. *18, 55, 56, 218.*

LUCKIESH, M. Comments on criteria of ease of reading. *J. exp. Psychol.*, 1946, 36, 180–2. *212.*

LUCKIESH, M., & MOSS, F. K. *The science of seeing*. New York: Van Norstand, 1937. *212.*

MACKWORTH, N. H. Researches in the measurement of human performance. MRC Special Report, No. 268. London: H.M.S.O., 1950. *113.*

MAHALANOBIS, P. C. On the generalised distance in statistics. *Proc. Nat. Inst. Sci.* (*India*), 1936, 12, 49–61. *299.*

MAHALANOBIS, P. C. Historical note on the D^2 statistic. *Sankhya*, 1948, 9, 237–45. *299.*

MARKS, M. R. A further investigation into the Kohler effect. *Amer. J. Psychol.*, 1949, 62, 62–74. *216.*

MARQUART, DOROTHY J. The satiational theory of figural after-effects. *J. gen. Psychol.*, 1954, 51, 83–91. *221.*

MARSHALL, W. H., & TALBOT, S. A. Recent evidence for neural mechanisms in vision leading to a general theory of sensory acuity. *Biol. Sympos.*, 1940, 7, 117-64. *218.*

MCCLELLAND, D. C. Factors influencing the time error in judgements of visual extent. *J. exp. Psychol.*, 1943, 33, 81–95. *125.*

MCGAUGHRAN, L. S. Predicting language behaviour from object sorting. *J. abnorm. soc. Psychol.*, 1954, 49, 183–95. *6.*

MCGAUGHRAN, L. S. Differences between schizophrenic and brain-damaged groups in conceptual aspects of object sorting. *J. abnorm. soc. Psychol.*, 1957, 54, 44–9. *6.*

MCGAUGHRAN, L. S., & MORAN, L. J. 'Conceptual level' vs. 'conceptual area' analysis of object-sorting behavior of schizophrenic and non-psychiatric groups. *J. abnorm. soc. Psychol.*, 1956, 52, 43–50. *6.*

MCLEOD, H. An experimental study of inheritance of introversion-extraversion. Unpublished Ph.D. Thesis. University of London, 1954. *193.*

MILLER, G. A. An auditory analogue of visual flicker. *Amer. Psychologist*, 1947, 2, 291. *131, 227, 228.*

MILLER, G. A. The perception of short bursts of noise. *J. Acoust. soc. Amer.*, 1948, 20, 160–70. *131, 227, 228.*

MILLER, G. A., & TAYLOR, W. G. The perception of repeated bursts of noise. *J. Acoust. soc. Amer.*, 1948, 20, 171–82. *131, 227, 228.*

MORAN, L. J. Vocabulary knowledge and usage among normal and schizophrenic subjects. *Psychol. Monogr.*, 1953, 67, No. 20 (Whole No. 370). *6, 7, 76.*

MOWBRAY, G. H., & GEBHARD, J. W. Differential sensitivity of the eye to intermittent white light. *Science*, 1955, 121, 173–5. *227.*

NACHMIAS, J. Figural after-effects in kinesthetic space. *Amer. J. Psychol.*, 1953, 66, 609–12. *217.*

NATHANSON, M., BERGMAN, D. S., & BENDER, M. Visual disturbances as the result of nystagmus on direct forward gaze. *Arch. Neurol. Psychiat.*, 1953, 69, 427–35. *212.*

NEEDHAM, J. G. The effect of the time interval upon the time error at different intensive levels. *J. exp. Psychol.*, 1935, 18, 530–43. *125.*

NICHOLS, E. G. *The relation between certain personality variables and the figural after-effect*. London: Ph.D. Thesis, 1956. *211, 217, 239.*

O'CONNOR, N., & RAWNSLEY, K. Incentives with paranoid and non-paranoid schizophrenics in a workshop. *Brit. J. Med. Psychol. 1959, 32, 133–143. 140, 165.*

OSGOOD, C. E. *Method and Theory in Experimental Psychology*. New York: Oxford Univ. Press, 1953. *125, 212.*

OSGOOD, C. E., & HEYER, A. W., jr. A new interpretation of figural after-effects. *Psychol. Rev.*, 1952, 59, 98–118. *218.*

OYAMA, T. Experimental studies of figural after-effects: I. Temporal factors. *Jap. J. Psychol.*, 1953, 23, 239–45. *216.*

PAGE, H. A., RAKITA, G., KAPLAN, H. K., & SMITH, N. B. Another application of the spiral after-effect in the determination of brain damage. *J. consult. Psychol.*, 1957, 21, 89–91. *205.*

PARDUCCI, A., & BROOKSHIRE, K. Figural after-effects with tachistoscopic presentation. *Amer. J. Psychol.*, 1956, 69, 635–9. *216.*

PAVLOV, I. P. *Conditioned reflexes.* London: Oxford University Press, 1927. *109, 193, 194.*

PAVLOV, I. P. *Conditioned Reflexes and Psychiatry* (W. H. Gantt, Trans.). London: Lawrence & Wishart, 1941. *161.*

PAYNE, R. W. *An Investigation into the Possibility of Defining 'Dissociation' as a Personality Trait by Means of Objective Tests.* Unpublished Ph.D. Thesis. Univ. of London Library, 1954. *4, 134.*

PAYNE, R. W. Cognitive abnormalities. In H. J. Eysenck (ed.), *Handbook of Abnormal Psychology*. London: Pitman (to appear 1960). *5, 17, 22, 30, 32, 35, 51, 53, 103.*

PAYNE, R. W., & HIRST, HEATHER L. Overinclusive thinking in a depressive and a control group. *J. consult. Psychol.*, 1957, 21, 186–88. *51, 53, 57.*

PAYNE, R. W., MATUSSEK, P., & GEORGE, E. I. An experimental study of schizophrenic thought disorder. *J. ment. Sci.*, 1959, 105, 627–52. *6, 7, 8–11, 23, 24, 27, 31, 33, 52, 54, 61, 62.*

PETRIE, ASENATH. *Personality and the frontal lobes.* London: Routledge & Kegan Paul, 1952. *123, 205.*

PETRIE, ASENATH, & LE BEAU, J. Psychological effects of selective frontal surgery including cingulectomy. *Proc. Vth Int. Congr. Neurol.*, Lisbon, 1953, 4, 392–5. *123, 205.*

PETRIE, ASENATH, & LE BEAU, J. A comparison of the personality changes after (1) prefrontal selective surgery for the relief of intractable pain and for the treatment of mental cases; (2) cingulectomy and topectomy. *J. ment. Sci.*, 1953, 99, 53–61. *123, 205.*

PHILBRICK, EMILY B. The validity of the spiral after-effect as a clinical tool for the diagnosis of organic brain pathology. *J. consult., Psychol.*, 1959, 23, 39–43. *205.*

PINILLOS, J. L., & BRENGELMANN, J. C. Bilderkennung als Personlichkeits-test. *Z. f. exp. u. angew. Psychol.*, 1953, 1, 480–500. *67.*

PONDER, E., & KENNEDY, W. P. On the act of blinking. *Quart. J. exp. Physiol.*, 1927–8, 18, 89. *212.*

POSTMAN, L. The time error in auditory perception. *Amer. J. Psychol.*, 1946, 59, 193–219. *125.*

PRENTICE, W. C. H. The relation of distance to the apparent size of figural after-effects. *Amer. J. Psychol.*, 1947, 60, 617–27. *217.*

PRENTICE, W. C. H. On the size of the figural after-effect with varying distances. *Amer. J. Psychol.*, 1950, 63, 589–93. *217.*

PRICE, A. C., & DEABLER, H. L. Diagnosis of organicity by means of spiral after-effects. *J. consult. Psychol.*, 1955, 19, 299–302. *123, 205.*

RAO, C. R., & SLATER, P. Multivariate analysis applied to differences between neurotic groups. *Brit. J. Stats. Psychol.*, 1949, 2, 17–29. *299.*

RAO, C. R. *Advanced statistical methods.* New York: John Wiley & Sons, 1952. *299.*

RAPAPORT, D., GILL, M., & SCHAFER, R. *Diagnostic Psychological Testing.* Chicago: Year Book Publishers, 1945. Vol. I. *6.*

RASHKIS, H. A. Three types of thinking disorder. *J. nerv. ment. Dis.*, 1947, 106, 650–70. *6.*

RASHKIS, H., CUSHMAN, T. F., & LANDIS, C. A new method for studying disorders of conceptual thinking. *J. abnorm. soc. Psychol.*, 1946, 41, 70–4. *6.*

RATLIFF, F., & RIGGS, L. A. Involuntary motions of the eye during monocular fixations. *J. exp. Psychol.*, 1950, 40, 687–701. *212.*

RAVEN, J. C. *The Mill Hill Vocabulary Scale.* London: Lewis, 1948. *9, 17.*

RAZRAN, G. H. S. Conditioned Responses in Children. *Am. Psychol.*, 1933, 23, 120. *167.*

RAZRAN, G. H. S. Conditioned responses: An experimental study and a theoretical analysis. *Arch. Psychol.*, 1935, 191, 1–24. *167, 168.*

RAZRAN, G. H. S. Attitudinal control of human conditioning. *J. Psychology*, 1936, 2, 327–37. *167.*

RAZRAN, G. H. S. Decremental and incremental effects of distracting stimuli upon the salivary C.R.S. of 24 Adult human subjects (inhibition and disinhibition), *J. Exp. Psychol.*, 1939, 24, 647–52. *167.*

RAZRAN, G. H. S. A direct laboratory comparison of pavlovion conditioning and traditional associative learning. *J. Abn. Soc. Psychol.*, 1955a, 51, 649. *167.*

RAZRAN, G. H. S. Partial Reinforcement of salivary C.R.S. in adult human subjects: preliminary study. *Psychol. Rep.*, 1955b, 1, 405–16. *167.*

RECHTSCHAFFEN, A. Neural satiation, reactive inhibition, and introversion-extraversion. *J. abn. soc. Psychol.*, 1958, 57, 283–91. *239.*

ROBY, T. B. An opinion on the construction of behavior theory. *Amer. Psychologist.*, 1959, 14, 129–134. *309.*

RODNICK, E. H., & SHAKOW, D. Set in the schizophrenic as measured by a composite reaction time index. *Amer. J. Psychiat.*, 1940, 97, 214–25. *109, 115.*

ROKEACH, M. Generalized mental rigidity as a factor in ethnocentrism. *J. abnorm. soc. Psychol.*, 1948, 43, 259–78. *56.*

SAGARA, M., & OYAMA, T. Experimental studies on figural after-effects in Japan. *Psychol. Bull.*, 1957, 54, 327–38. *216.*

SHAPIRO, M. B. Experimental studies of a perceptual anomaly. III. The testing of an explanatory theory. *J. ment. Sci.*, 1953, 99, 394–409. *205.*

SHAPIRO, M. B. A preliminary investigation of the effects of continuous stimulation on the perception of apparent movement. *Brit. J. Psychol.*, 1954, 45, 58–67. *217.*

SIGAL, J. J., STAR, K. H., & FRANKS, C. M. Hysterics and dysthymics as criterion groups in the study of introversion-extraversion. *J. Abnorm. soc. Psychol.*, 1958a, 57, 143–8. *112, 135, 166.*

SIGAL, J. J., STAR, K. H., & FRANKS, C. M. Hysterics and dysthymics as criterion groups in the measure of introversion-extraversion: a rejoinder to Eysenck's reply. *J. abnorm. soc. Psychol.*, 1958*b*, 57, 381–2. *112, 135, 166.*

SLATER, P. The factor analysis of matrices of negative correlations. *Brit. J. Stats. Psychol.*, 1953, 6, 101–6. *288.*

SMITH, K. R. The satiational theory of figural after-effects. *Psychol. Rev.*, 1948, 61, 282–6. *218.*

SMITH, K. R. The statistical theory of figural after-effects. *Psychol. Rev.*, 1952, 59, 401–2. *218.*

SPENCE, K. W., & FARBER, J. E. Conditioning and extinction as a function of anxiety. *J. Exp. Psychol.*, 1953, 45, 116–19. *161.*

SPENCE, K. W., & TAYLOR, JANET, A. Anxiety and the strength of the U.C.S. as determiners of the amount of eyelid conditioning. *J. Exp. Psychol.*, 1951, 42, 183–8. *161.*

SPENCE, K. W., & TAYLOR, JANET A. The relation of conditional response strength to anxiety in normal, neurotic and psychotic subjects. *J. Exp. Psychol.*, 1952, 45, 265–72. *161.*

STANDLEE, L. S. The Archimedes negative after-effect as an indication of memory impairment. *J. consult. Psychol.*, 1953, 17, 317. *205.*

STAR, K. H. *An experimental study of 'reactive inhibition' and its relations to certain personality traits*. Ph.D. thesis, Univ. of London, 1957. *102.*

STEVENS, S. S., & DAVIES, H. *Hearing, its psychology and physiology*. New York: Wiley, 1938. *227.*

STORMS, L. H. Discrepancies between factor analysis and multivariate discrimination among groups as applied to personality theory. *J. ment. Sci.*, 1958, 104, 713–21. *271, 293.*

SUMMERFIELD, A., & MILLER, K. M. Visual illusions and figural after-effect, with and without fixation. *Quart: J. exp. Psychol.*, 1955, 7, 149–58. *211.*

SUTHERLAND, N. S. Figural after-effects, retinal size and apparent size. *Quart. J. exp. Psychol.*, 1954, 6, 35–44. *217.*

SYMMES, D., CHAPMAN, L. F., & HALSTEAD, W. C. The fusion of intermittent white noise. *J. Acoust. soc. Am.*, 1955, 27, 470–3. *131, 227, 228.*

THALMAN, W. A. The after-effect of seen movement when the whole visual field is filled by a moving stimulus. *Amer. J. Psychol.*, 1921, 32, 429–40. *203, 204.*

THOMSON, G. H. *Factorial analysis of human ability*. London: Univer. London Press, 1950. *250, 251.*

THURSTONE, L. L. A factorial study of perception. *Chicago: Univ. Chicago Press*, 1943. *194, 204.*

THURSTONE, L. L. *Multiple Factor Analysis*. Chicago: Univ. of Chicago, 1947. 83.

TIZARD, J., & VENABLES, P. H. The influence of extraneous stimulation on the reaction time of schizophrenics. *Brit. J. Psychol.*, 1957, 48, 299–305. *109, 115.*

U.S. EMPLOYMENT SERVICE. *Guide to the Use of General Aptitude Test Battery.* Washington, D.C.: U.S. Dept. of Labor, 1955. *69.*

VAN DER HURST. 1924. Quoted by Eysenck, 1952a. *4.*

VAN DER MERWE, A. B. The diagnostic value of peripheral vasomotor reactions in the psychoneuroses. *Psychosom. Med.*, 1948, 10, 357–54. *136.*

VAUCOULEURS, G. DE. Discovery of the Universe. *London: Faber*, 1957. *315.*

VENABLES, P. H. Factors in the motor behaviour of functional psychotics. *J. abnorm. soc. Psychol.*, 1959, 58, 153. *109, 115, 128.*

VENABLES, P. H., & TIZARD, J. Performance of functional psychotics on a repetitive task. *J. abnorm. soc. Psychol.*, 1956a, 53, 23–6. *109.*

VENABLES, P. H., & TIZARD, J. Paradoxical effects in the reaction time of schizophrenics. *J. abnorm. soc. Psychol.*, 1956b, 53, 220–4. *109.*

VIGOTSKY, I. S. Thought in schizophrenia. *Arch. Neurol. Psychiat., Chicago*, 1934, 31, 1063–77. *6, 34.*

WALLS, G. L. Interocular transfer of after-images. *Amer. J. Optometry*, 1953, 30, 57–64. *203.*

WALTHALL, W. T. The Kohler effect. *Amer. J. Psychol.*, 1946, 59, 152–5. *221.*

WECHSLER, D. *The Measurement of Adult Intelligence.* William & Williams Co., Baltimore, 1944. *18, 65.*

WEGROCKI, H. Generalizing ability in schizophrenia: An inquiry into the disorders of problem thinking in schizophrenia. *Arch. Psychol.*, 1940, 36, No. 254. *7.*

WENGER, M. A. Studies of autonomic balance in Army Air-Force personnel. *Compar. Psychol., Monographs*, 1948, 101, *Vol. 19*, No. 4, *173, 174.*

WERTHEIMER, M. The differential stability of schizophrenic and normal subjects: A test of deduction from the theory of figural after-effects. *J. gen. Psychol.*, 1954, 51, 291–9. *218.*

WERTHEIMER, M. Figural after-effects as a measure of metabolic efficiency. *J. Pers.* 1955, 24, 56–73. *218.*

WERTHEIMER, M., & WERTHEIMER, NANCY. A metabolic interpretation of individual differences in figural after-effects. *Psychol. Rev.*, 1954, 61, 279–80. *218.*

WERTHEIMER, M., & WERTHEIMER, NANCY. The effect of experimentally induced changes in metabolism on perceptual measures of metabolic efficiency. *Percept. Mot. Skills*, 1955, 5, 173–6. *218.*

WESTHEIMER, G., & CONOVER, D. W. Smooth eye movements in the absence of a moving visual stimulus. *J. exp. Psychol.*, 1954, 47, 283–4. *212.*

WHITE, MARY A. A study of schizophrenic language. *J. abnorm. soc. Psychol.*, 1949, 44, 61–74. *8.*

WILKS, S. S. Certain generalisations in the analysis of variance. *Biometrika*, 1932, 24, 471–8. *299.*

WILLETT, R. A. *A Handbook of Experimental abnormal Psychology* (ed. Eysenck, H. J.). Pitman, London (to appear). *166.*

WOHLGEMUTH, A. On the after-effect of seen movement. *Brit. J. Psychol., Monogr. Suppl.*, 1911, 1. *203.*

WOOD, C. L., & BITTERMAN, M. E. Blinking as a measure of effort in visual work. *Amer. J. Psychol.*, 1950, 63, 584–8. *212.*

WOODROW, H. Individual differences in the reproduction of temporal intervals. *Amer. J. Psychol.*, 1933, 45, 371–81. *129.*

WYATT, S., & LANGDON, J. N. Fatigue and boredom in repetitive work. *MRC report, No. 77*, Industrial Health Research Board. London: H.M.S.O., 1937. *129.*

SUBJECT INDEX

abstract thought, in schizophrenia, 6–7, 9, 27–66 (see also *concreteness*)
accuracy, of problem solving, 5, 10, 19, 59–63
adaptation, 215, 216
affect, inappropriate, 4
annoyances test, 273
archimedes spiral, 195, 197, 212, 214, 218
archetypes, 308
arousal, 136–7, 154
atomic vibration, 314
attention, dissociation of, 3–5
auditory flutter fusion, 131–2, 195, 196, 227–31, 232, 236
auditory flutter fusion, and extraversion, 132, 146, 152
auditory flutter fusion, and intelligence, 145
auditory flutter fusion, in neurotics, 132
auditory flutter fusion, in psychotics, 132

Babcock-Levy motor speed tests, 18, 23–26, 83, 84, 98
blinking, 212, 213
body build index, 273
body sway test, 133, 150
brain damage, 214
Brengelmann picture recognition test, 17, 66–8, 84, 98, 132–3, 151

Cattell's CPF, 197, 210, 211, 215, 231, 232
circulation, of the blood, 306
co-incident movement, 213
comparison of circles, 219
concentration, see *distraction*
concept formation, in schizophrenia, 6–9, 10–11, 13, 27–66
concept formation, in depression, 27–66
concept formation, in neurosis, 9, 10, 27–66 (see also *concreteness*, and *overinclusive thinking*)
concreteness, in schizophrenia, 6–7, 9, 10–11, 13, 26–29
concreteness and intelligence, 100–1
concreteness, Feldman-Drasgo test of, 10
concreteness, Goldstein-Scheerer tests of, 6, 10, 18, 26–7, 84, 93, 95, 101
concreteness, Payne object classification test of, 10, 11, 18, 32–5, 84, 95, 98, 101
concreteness, Proverbs test of, 10–11, 27–9, 84, 93, 95, 96, 98, 101
concreteness, Shaw test of, 42–51, 84, 93, 95, 96, 98, 101
concreteness, Vigotsky test of, 6
conditionability, 311
conditioned inhibition, 136
conditioned inhibition and vigilance, 114
conditioned inhibition in psychotics, 139–40
conditioning, eyeblink, 160–7
conditioning, measurement of, 311
conditioning, salivary, 167–74, 236
conditioning, spatial, 183–91, 234, 236
conditioning, verbal, 174–82, 234, 237
constitution, in schizophrenia and manic-depressive psychosis, 3–5
Contact Personality Inventory, 159, 163, 164, 170, 173, 178, 179, 187–90
continued stimulation, 194
contrast, 221
cortical conductivity, 217, 218
critical level, 136

dark adaptation, 195, 196, 200–2, 225, 227, 231, 232, 236
delusions, 4, 103–4
depression, and thought disorder, 3–104
depression, and intellectual slowness, 5–6, 19, 97–100
depression, and overinclusion, 11, 29–66, 81, 97
depression, and concreteness, 26–29, 100–1
discreteness, 228
discrimination learning, and concept formation, 8–9
discrimination learning, and overinclusion, 8–9, 55–9

disinhibition, 114, 118
disinhibition, and extraversion, 146
dissociation, 3–5
distraction, in psychotics, 6, 12, 23, 97–98, 101–3
double D test, 219
drawing design test, 19, 62–5, 84, 98
drive, and inhibition, 136–7
drive, and neuroticism, 135
drive, and reminiscence, 136, 137, 146, 150
drive, factor of, 146–7, 150
drive, in hysterics and dysthymics, 135–7, 153
drugs, 204–5
dynamometer test, 18, 72–3, 84

educational attainment, of psychotics and neurotics, 16–17
Epstein overinclusion test, 18, 51–5, 84, 95, 98, 101
error, in problem solving, see *accuracy*
excitation, 194, 215, 222
excitation-inhibition balance, shift in hysterics and dysthymics, 136–7, 147, 151, 154
experimentum crucis, 306
expressive movement in psychotics and neurotics, 71–2
extraversion, 157, 158, 160, 164, 167, 177, 179, 183, 190, 194, 201, 205, 209, 211, 215, 218, 224, 225, 226, 229, 231, 232, 233
extraversion and AFF, 132, 146, 152
extraversion, and inhibition, 108, 112, 119, 135, 151
extraversion, and proverbs test, 31
extraversion, and reminiscence, 122, 146, 151
extraversion, and time error, 128, 146, 151
extraversion, and time judgement, 130, 131, 146, 152
extraversion, and vigilance, 114, 117–18, 146, 151
extraversion, in hysterics, 112, 135
extraversion, factor of, 146, 150–1, 235
eye blink conditioning, 236
eye movements, 212, 214, 219

factor analysis, 196, 234
factors, 194
figural after-effect, 195, 196–7, 207, 211, 215–27, 231, 237
figural after-effect, and overinclusion, 11
fixation, 195, 196, 211–15, 217, 225
fixation, failure to maintain, 237
food aversions, 273
fragmentation of mental life, 3–5
frequency matching, 174, 176, 177

gas laws, 314
genetic determinants, 193
Goldstein-Scheerer Colour-Form test, 18, 26–7, 84, 93, 101
Goldstein-Scheerer Object Sorting test, 18, 26–7, 32–5, 84, 93, 95, 101

hysteria, 194, 205, 209, 217
hysteria, and extraversion, 313
hypothetico-deductive method, 303

information theory, and overinclusion, 9, 11
inhibition, 194, 195, 202, 204, 205, 228, 231, 232, 233
inhibition, and overinclusion, 8–9, 11
inhibition theories, 193, 308
inhibitory tendencies, 157, 158, 160, 167
integration-disintegration of mental processes, 3–5
intelligence, 197
intelligence, and concreteness, 100–1
intelligence, and overinclusion, 94–5
intelligence, and vocabulary, 9, 13, 16, 17, 100
intelligence, factor of, 81–92, 100–1
intelligence, matrices test of, 9, 100
intelligence, Nufferno test of, 9, 20–1, 100
intelligence, timed tests of, 5–6
inter-ocular transfer, 221, 225
introversion, see *extraversion*
introversion, factor of, 236, 237

J.N.D., 228

Lagrange multiplier, 251
leptomorphic physique, and psychosis, 3–5, 11–12
Leiter-Partington Pathways test, 19, 59–62, 84, 93, 98
Luchins Einstellung rigidity test, 18, 55–59, 82, 84, 94, 95

manic-depressive psychosis, 3–5
manic-depressive psychosis, and psychoticism, 11–12 (see also *depression*)

manifest anxiety, 161
mass-luminosity principle, 314
Maudsley Medical Questionnaire, 273
Maudsley Personality Inventory (M.P.I.), 111–12, 146, 150, 159, 163, 166–7, 170, 173, 178, 179, 187–190, 196, 197, 209, 210, 215, 231, 232, 235, 237
memory impairment, 205
Mill Hill Vocabulary Test, 16–17, 84, 100, 273
Minnesota Multiphasic Personality Inventory (M.M.P.I.), 111–12, 146, 150, 159, 163, 164, 170, 173, 178, 179, 187–190, 197, 215, 231, 235, 237
motivation, 232, 233
motor tests, in psychosis, 4, 10, 12, 23–6, 68–71, 97–100

neologisms, in schizophrenia, 7, 54–5, 101
neurosis, and concreteness, 26–9, 100–1
neurosis, and overinclusion, 8–11, 29–66, 81–100
neurosis, and psychoticism, 11–12, 81–100
neurosis. and retardation, 8–11, 81–100
neuroticism, 161, 164, 167, 179, 190
neutrino, 309
N.I.I.P. test 90A, 159, 181, 199
Nufferno paper test, 273
Nufferno speed and level tests, 9, 18, 19–23, 66, 84, 96, 98, 100

occupational ratings, in neurosis and psychosis, 16–17
Oedipus complex, 308
organicity, 205
overinclusive thinking, and discrimination learning, 8–9
overinclusive thinking, and Einstellung rigidity, 55–9, 84, 94, 95
overinclusive thinking, and inductive reasoning, 65–6, 84
overinclusive thinking, and information theory, 9
overinclusive thinking, and intelligence, 94–5
overinclusive thinking, and retardation 10, 25, 97–100
overinclusive thinking, and schizophrenia, 7–11, 12–13, 80
overinclusive thinking, and stimulus generalization, 9
overinclusive thinking, Cameron's theory, 7–8
overinclusive thinking, drawing design test, 62–5, 84
overinclusive thinking, Epstein test, 7, 11, 18, 51–5, 84, 95, 98, 101
overinclusive thinking, factor of, 81–97
overinclusive thinking, Goldstein object sorting test, 11, 18, 32–5, 84, 95, 101
overinclusive thinking, Pathways test, 11, 59–62, 84, 93
overinclusive thinking, Payne, Matussek and George study, 8–11
overinclusive thinking, Payne object classification test, 11, 17, 35–42, 84, 95, 98, 101
overinclusive thinking, proverbs test, 11, 29–32, 84, 93, 95, 96, 98, 101
overinclusive thinking, Shaw test, 42–51, 84, 93, 95, 96, 98, 101

Payne object classification test, 11, 17, 35–42, 82, 84, 95, 98, 101
perception, accuracy of, 62–5
perception, speed of, 12, 66–8, 80
perception, tests in psychosis, 4
perceptual abnormalities, 193
persistence, 10, 22–3
persistence test, 273
personality and physique, 3–5, 11–12
physique, leptomorphic-pyknic, 3–5, 11–12
Porteus maze test, 273
principal component analysis, 247
Progressive Matrices test, 273
protective inhibition, 109, 115, 126, 128
proverbs, Benjamin test, 7, 10, 19, 27–32, 84, 93, 95, 98, 101
proverbs, Gorham test, 7
P.S.Es, 220
psychosis, 3–9, 11–12
psychosis, thought disorder in, 3–104
psychoticism, factor, 11–12, 97–100
psychoticism, tests, 66–79, 80, 81–92
pursuit rotor, 119–22
pursuit rotor, and drive, 136, 147, 150
pursuit rotor, performance in neurotics, 121
pursuit rotor, performance in psychotics, 121, 139–41
pursuit rotor, reminiscence, 122, 136, 137, 140–1
pyknic physique, and psychosis, 3–5, 11–12

questionnaires, 198–9
questionnaire neuroticism, 146, 150

reactive inhibition, and drive, 109, 136
reactive inhibition, and extraversion, 108, 112, 135, 151
reactive inhibition, and overinclusive thinking, 11
reactive inhibition, and reminiscence, 119
reactive inhibition, and satiation, 122–3
reactive inhibition, and vigilance, 114
reactive inhibition, decay rate in psychotics, 128, 138–41, 153
reactive inhibition, in hysterics and schizophrenics, 141, 154
relativity theory, 314
reminiscence, 122, 204
reminiscence, and drive, 136, 137, 146, 150
reminiscence, and extraversion, 122, 146, 151
reminiscence, and habit strength, 137
reminiscence in neurotics, 122
reminiscence in psychotics, 122, 140–1
retardation, Babcock tests of, 8–11, 18, 23–6, 83, 84, 98
retardation, factor of, 81–92, 97–100
retardation of thought, 5–6, 9–10, 19–26, 66–71, 80
retardation, Nufferno tests, 8–11, 18, 19–23, 66, 84, 96, 98, 100
rigidity, 10

salivary conditioning, 167–174, 236
satiation, 122–3, 194, 195, 216, 232, 233, 305
scalars, 251
schizophrenia, Archimedes spiral in, 124, 138
schizophrenia, protective inhibition in, 109, 115, 126, 128
schizophrenia, pursuit rotor performances in, 121, 139–41
schizophrenia, and reminiscence, 122, 140–1
schizophrenia, straight-line learning curve, 139, 140
schizophrenia, warm-up, 139–40
schizophrenia, vigilance in, 116
schizophrenic thought disorder, 3–9
schizothymia-cyclothymia, 3–5, 11–12
set, and overinclusion, 54–65
sex humour test, 273
Shaw test, 18, 19, 42–51, 82, 84, 93, 95, 96, 98, 101
slowness, Babcock-Levy tests of, 8–11, 18, 23–26, 83, 84, 98
slowness, intellectual, 5–6, 9–10, 19–26, 66–71, 80, 81–92, 97–100
slowness, Nufferno tests of, 8–11, 19–23, 66, 84, 96, 98
sorting tests, 133–4
sorting tests, and intelligence, 145, 149
sorting tests, Goldstein-Scheerer, 6, 10, 26–7, 32–5, 84, 93, 95, 101
sorting tests, Payne object classification test, 10, 11, 17, 35–42, 82, 84, 95, 98, 101
sorting tests, Shaw test, 18, 19, 42–51, 82, 84, 93, 95, 96, 98, 101
sorting tests, Vigotsky test, 6
sound discrimination test, see *auditory flutter fusion*
Spaltungsfähigkeit, see *dissociation*
spatial conditioning, 183–91, 234, 236
speed of thinking, 5–6, 9, 12, 19–26, 66–71, 80, 81–92, 97–100 (see also *retardation*)
spiral, 122–5
spiral after-effects, 203–11, 227, 231, 236
spiral, and drive, 146, 147
spiral, and excitation-inhibition balance, 148
spiral, and extraversion, 124, 146, 151
spiral, derived scores, 207
spiral, in neurotics, 124
spiral, in psychotics, 124, 138
static ataxia, 18, 73–6, 84, 273
statistical theory, 218
S.T.D.C.R. scales, 164, 165
stimulus generalization, and overinclusion, 9
stress-gain, 9
suggestibility test, 273

theory, in psychology, 303
thought disorder, in psychotics, 3–104
three squares test, 18, 68–9, 84
time error, 125–8
time error, and extraversion, 128, 146, 151
time error, in neurotics, 127–8
time error, in psychotics, 127–8
time judgement, 228–31
time judgement, and extraversion, 130–1, 146, 152
time judgement, and intelligence, 145

time judgement, in neurotics, 129–30
time judgement, in psychotics, 129–30

USES manual and finger dexterity tests, 19, 69–71, 84, 96, 98

verbal conditioning. 174–82, 234, 237
vigilance, 113–19
vigilance, and drive, 136, 147, 150
vigilance, and extraversion, 114, 117–18, 146, 151
vigilance, and intelligence, 145
vigilance, in neurotics, 116–17
vigilance, in psychotics, 116–17
vocabulary, and intelligence, 9, 13, 16, 17
vocabulary, Mill Hill test of, 13, 16, 17, 84, 100

waves test, 18, 71–2, 84
Wechsler arithmetic test, 18, 65–6, 84, 100
Wechsler picture arrangement test, 18, 65–6, 84, 96, 100
word association test, 18, 76–9, 84, 93, 96

PSYL. RESEARCH

S.C.E.R.T., West Bengal
Library
Calcutta
25/3, B. C.